UNIVERSITY OF CALIFORNIA
Santa Barbara

An Analysis of the Efficacy and Treatment Trajectory of the START Social Skills Group for Adolescents with Autism Spectrum Disorder

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy in
Counseling, Clinical, and School Psychology

by

Amber R. Miller

Committee in charge:

Professor Ty Vernon, Chair

Professor Robert Koegel

Professor Matthew Quirk

September 2016

The dissertation of Amber R. Miller is approved.

_____

Robert Koegel

_____

Matthew Quirk

_____

Ty Vernon, Committee Chair

June 2015

ACKNOWLEDGEMENTS

Thank you to the many people who have supported me on this journey…

TMM, you have always inspired me with your dedication to education and science. It's by your side that I learned to be analytical and to persist in the search for the truth. Thank you for teaching me to be a lifelong learner.

RSM, you have always believed in me. Your excitement about my work and my life encourages me and reminds me to believe in myself. You've taught me to trust in what my life has in store for me while doing what I can to make the world a better place.

PJB, you make me a better person every day. You've been the best long-distance support I ever could have imagined and I'm still sane because of you.

RCM, no matter how different our lives become, I always know I have a soul sister in you. You remind me to keep a balance in my life and to be true to myself.

TWV, I couldn't have asked for a better advisor. Your openness, trust, and guidance allowed me to actually enjoy graduate school. Thank you for both supporting me and pushing me to improve.

BK & LK, your work has inspired me ever since I first learned about Autism. It has been a privilege to work with and learn from you both. You've given me so many amazing opportunities and you've brought together an incredible group of people to learn with, thank you.

KA, JB, AN, AV, and others, this lab has felt like family. Your research and clinical skills amaze me, someday you will all be leaders in the field. I'll never forget our happy hours, hot yoga, races, and travels. You made grad school fun!

JE, LT, & AC, you were literally there for me every day. You will never be replaced. Thank goodness for Craig!

VITA OF AMBER R. MILLER

June 2015

EDUCATION

Anticipated Ph.D. in Counseling, Clinical, and School Psychology — June 2016
Clinical Psychology emphasis
University of California, Santa Barbara
Dissertation Title: *An Analysis of the Efficacy and Treatment Trajectory of the START Social Skills Group for Adolescents with Autism Spectrum Disorder*
Advisors: Dr. Ty Vernon and Dr. Robert Koegel

M.A. in Counseling Psychology — June 2013
University of California, Santa Barbara
Thesis Title: *An Analysis of a Video-Feedback Intervention for Adults with Autism Using a Multiple Baseline Experimental Design*

B.S. in Psychology, with an emphasis in Biology — June 2007
Minor in Communications
University of California, Davis
Graduated with Honors

RESEARCH AND TEACHING POSITIONS

University of California, Santa Barbara
Department of Counseling, Clinical, and School Psychology
Graduate Student Researcher, Koegel Autism Center — 2011 – 2015
Teaching Assistant — 2014 – 2015

University of California, San Francisco
Department of Child and Adolescent Psychiatry
Autism and Neurodevelopment Program
Clinical Research Coordinator — 2009 – 2011

University of California, Davis Medical Center
Department of Psychiatry
Clinical Research Coordinator — 2007 – 2009

Charles B. and Linda C. Schaffer, M.D., Inc., Sacramento, CA
Research Assistant — 2006 – 2009

PEER-REVIEWED PUBLICATIONS

Miller, A. R., Vernon, T. W., Wu, V., & Russo, K. (2014). Social skill group interventions for adolescents with autism spectrum disorders: A systematic review. *Review Journal of Autism and Developmental Disorders, 1* (4), 254-265. doi: 10.1007/s40489-014-0017-6

Koegel, L. K., Koegel, R. L., Miller, A. R., & Detar, W. J. (2014). Issues and interventions for autism spectrum disorders during adolescence and beyond. In F. R. Volkmar, R. Paul, & K. A. Pelphrey (Eds.), *Handbook of Autism and Pervasive Developmental Disorders* (4th ed., Vol. 1, pp. 176–190). Hoboken, NJ: Wiley.

Maddock, R. J., Buonocore, M. H., Miller, A. R., Yoon, J. H., Soosman, S. K., & Unruh, A. M. (2013). Abnormal activity-dependent brain lactate and glx responses in panic disorder. *Biological Psychiatry, 73* (11), 1111-1119. doi: 10.1016/j.biopsych.2012.12.015.

Schaffer, L. C., Schaffer, C. B., Miller, A. R., Manley, J. L., Piekut, J. A., & Nordhal, T. E. (2013). An open trial of pregabalin as an acute and maintenance adjunctive treatment for outpatients with treatment resistant bipolar disorder. *Journal of Affective Disorders,147,* 407-410. doi: 10.1016/j.jad.2012.09.005.

Schaffer, C. B., Schaffer, L. C., Miller, A. R., Hang, E., & Nordhal, T. E. (2011). Efficacy and safety of nonbenzodiazepine hypnotics for chronic insomnia in patients with bipolar disorder. *Journal of Affective Disorders, 128* (3), 305-308. doi: 10.1016/j.jad.2010.07.018

Schaffer, L. C., Schaffer, C. B., Hunter, S. L., & Miller, A. R. (2010). Psychiatric reactions to isotretinoin in patients with bipolar disorder. *Journal of Affective Disorders, 122* (3), 306-308. doi: 10.1016/j.jad.2009.09.005

HONORS & AWARDS

Organization for Autism Research (OAR)	
Applied Research Grant: $29,741	2014 – 2015
University of California, Santa Barbara	
Graduate Fellowship	2011 – 2015
Dean's Fellowship	2012 – 2013
Outstanding Research Award	2013
University of California, Davis	
Dean's Honor List	2003 – 2007
National Dean's Honor List	2007
Phi Beta Kappa National Honor Society	2007
Golden Key International Honor Society	2005 – 2007
National Society of Collegiate Scholars	2003
Phi Kappa Phi Honor Society	2003

ABSTRACT

An Analysis of the Efficacy and Treatment Trajectory of the START Social Skills Group for Adolescents with Autism Spectrum Disorder

by

Amber R. Miller

Adolescence can be a time of new and complex social challenges and individuals with Autism Spectrum Disorder (ASD) are in need of evidence-based interventions to bolster their social skills during this critical developmental period. While a number of social skill group interventions have been developed, many of these are in need of further research evidence of their effectiveness. A literature review was conducted and identified a number of strengths and weaknesses in the currently available interventions. The START group was designed to incorporate those strengths and improve upon those weaknesses. A previous pilot study of the START group indicated that it may be an effective intervention for this population. The current study expanded on this pilot study and investigated several important questions about the START group, including its effectiveness when compared to a waitlist control group and the trajectories of improvement of participants. Nineteen adolescents with ASD participated, and a number of outcome measures were utilized to examine social change, including parent report measures, participant self-report measures, behavioral observations, and subjective ratings. The analysis comparing the treatment and control groups resulted in large effect sizes on both the primary parent and participant

measures. Other measures also exhibited small to large effect sizes. These results were not found to be statistically significant, which may be attributed to the small sample size. However, the relatively large effect sizes found indicate the clinical and social significance of this intervention. The analysis of the treatment trajectories of participants indicated that most of the outcome measures exhibited statistically significant improvement after only ten weeks of intervention. After this initial gain, the primary parent report measure continued to exhibit statistically significant improvement throughout the duration of the treatment. Other measures indicated positive but non-significant trends toward continued improvement. Possible explanations for this treatment trajectory are discussed. The clinical implications of these findings are explored, as are future research directions.

TABLE OF CONTENTS

Introduction

Autism Spectrum Disorder (ASD) is a pervasive neurodevelopmental disorder characterized by delays or deficits in two core areas: social communication and restricted interests/repetitive behaviors (American Psychiatric Association, 2013). The prevalence of ASD has been rising in recent decades and is currently being estimated at 1 in every 68 children (CDC, 2014).

The ever-increasing rates of ASD have led to a great demand for treatments, and the past several decades have seen an explosion of new intervention strategies (Simpson et al., 2005), one of which is social skills training. Social skills intervention is clearly imperative to address the core area of socialization difficulties in individuals with ASD. Not surprisingly, social skills interventions have increased in popularity and research has shown that these types of interventions have emerging empirical evidence for their effectiveness (National Autism Center, 2009). Many of these social skills interventions have been held in group formats, as this type of setting may be an ideal way to give individuals the opportunity to practice newly learned skills with their peers (Barry et al., 2003).

However, there is limited research available on the effectiveness of these types of interventions for adolescents. Research has tended to focus primarily on early intervention and has largely ignored the adolescent population until recently (Williams-White, Koenig, & Scahill, 2007). While early intervention is critical, we are now faced with a large number of teens who are still in need of services. In particular, higher-functioning adolescents with ASD are especially susceptible to the dearth of available intervention services (Williams-White, Koenig, & Scahill, 2007). These types of children tend to "fall through the cracks"

because their relatively mild symptoms are overlooked in favor of treating the more impaired children with autism.

Adolescents with ASD

Research on the developmental progression of individuals with ASD has found that they continue to exhibit difficulties with social interactions and engagement through adolescence. Sigman and Ruskin (1999) even go so far as to assert that individuals with ASD will likely be as severely affected by the core symptoms of autism in adolescence as in early childhood. Recently, Schall and McDonough (2010) summarized the literature on the characteristics of adolescents with ASD and noted that while adolescents with ASD generally show improvements in communication, they continue to exhibit distinct impairment in *social* communication. Clearly, intervention is not something that can be terminated after childhood, but is indicated throughout adolescence. Seltzer and colleagues (2003) studied a large group of adolescents and adults with ASD and found that, while significant improvements occurred in several domains over time, the two areas showing the least improvement with age were the presence of friendships and the presence of circumscribed interests. All of this research points to enduring impairments that are not resolved prior to adolescence.

In addition to the enduring difficulties experienced by individuals with ASD, several new challenges are presented as they enter adolescence. One new challenge is exposure to increasingly complex social situations (Green, Gilchrist, Burton, & Cox, 2000). As children mature, their friendships tend to become more verbally-based, clearly something with which individuals with ASD struggle. Adolescents also begin to experience greater independence from their parents and venture into the new social territories of dating and employment.

Each of these domains presents potential new difficulties for adolescents with ASD, and intervention at this stage is imperative to improve chances at success in these areas.

Social Skills Groups for Adolescents with ASD

Research on social skills interventions has dramatically increased in the past two decades and "tentative" support has been established for the effectiveness of these interventions in group formats (Reichow & Volkmar, 2010). In recent years many new social skills groups have been developed by researchers across the US (and the world), but many of these studies have focused on providing services to school-aged children with ASD. While social skills groups for this age range are considered an established evidence based practice (Reichow & Volkmar, 2010), group treatments for adolescents lag behind in terms of their empirical support.

A review of the literature identified 44 articles that directly targeted social skills in adolescents with ASD in a group format[*]. While each identified treatment utilized a group format, there was significant heterogeneity among the implemented procedures and evaluative methods. A number of dimensions served as a means for distinguishing between identified group interventions, including theoretical basis, delivery format, specifically targeted skills, treatment individualization/customization, generalization and maintenance considerations, duration and intensity of treatment, participants per group, inclusion of typically developing peers, the cognitive range of the participants, and intervention context. In addition, differences were noted in the study designs and the methods of assessing change presented in the articles. Each of these important domains is addressed below and ultimately

[*] The information contained in this review is also represented in a recently published article in the Review Journal of Autism and Developmental Disabilities (Miller, Vernon, Wu, & Russo, 2014).

influenced both the development of a new adolescent social skills group as well as the methods used to evaluate it in the current research study.

Study designs. The reviewed investigations varied considerably with regard to research designs. While single-case designs are frequently used in autism research, only a handful of investigations (e.g. Dotson, Leaf, Sheldon, & Sherman, 2010; Mitchell, Regehr, Reaume, & Feldman, 2010; Webb, Miller, Pierce, Strawser, & Jones, 2004) used this type of design, which may be attributed to difficulties adopting single-case methods to a group intervention context. Some articles assessed program feasibility characteristics, such as fidelity and consumer satisfaction (e.g. White, Albano, et al., 2010), while others used qualitative methods, such as categorizing participant and parent responses to socially relevant inquiries and questionnaires (Fullerton & Coyne, 1999; Rose & Anketell, 2009). A large proportion of studies (39%) used a pre-post design to assess treatment effects without a specified control group (e.g. Herbrecht et al., 2009; MacKay, Knott, & Dunlop, 2007; Stichter et al., 2010; Tse, Strulovitch, Tagalakis, Meng, & Fombonne, 2007). In contrast, a very small number (6 of the 44) used a randomized controlled trial (RCT; Laugeson, Frankel, Gantman, Dillon, & Mogil, 2012; Laugeson, Frankel, Mogil, & Dillon, 2009; White et al., 2013) to evaluate treatment effects. Many researchers who opted to study efficacy using uncontrolled or quasi-controlled methods indicated their intention of implementing an RCT after the establishment of initial efficacy and feasibility, following NIMH current research evaluation recommendations (Smith et al., 2007). Additional implementations of carefully controlled RCTs will likely be required in the future to provide clearer evidence of treatment acceptability and effectiveness.

Setting. Many of the studies examined in this review took place in clinic settings (63%), often associated with large research universities. However, some of the described investigations evaluated adolescent social skills groups in school settings (Minihan, Kinsella, & Honan, 2011; Williams, 1989) and other settings, such as on horseback (Gabriels et al., 2012), at a musical theater company (Corbett et al., 2011), in computer workshops (Wright et al., 2011), and in summer camps (Lopata, Thomeer, Volker, Nida, & Lee, 2008). While most studies reviewed took place in the United States, several were conducted in other countries, such as Germany (Herbrecht et al., 2009), Scotland (MacKay et al., 2007), Ireland (Minihan et al., 2011; Rose & Anketell, 2009), Italy (Valenti, Cerbo, Masedu, De Caris, & Sorge, 2010), and South Korea (Shin, Koh, & Yeo, 2012).

Number of participants per group. There is an established trend of including groups of 4 to 6 participants within a given social skills group paradigm. While not all publications reviewed noted the exact number of participants within each individual group, most fell within this range. This number of participants may be perceived as the ideal balance of group camaraderie and ample opportunities for individual attention, but no studies to date have evaluated the effect of participant numbers on treatment effectiveness.

The PEERS group falls on the higher end of the range with eight to ten participants in each group (Laugeson et al., 2009). The PEERS intervention yielded significant treatment gains immediately following intervention as well as during a 14-week follow-up assessment. These findings suggest that while groups of 4 to 6 participants seem to be favored among researchers, it is possible to deliver group interventions to larger numbers of adolescents. However, another factor to consider during group size determinations may be the teaching strategy that will be employed. Specifically, didactic lessons may be easier to

scale up for larger groups. Other groups that focus primarily on experiential or activity-based strategies may find that smaller groups allow participants more opportunities for practice and feedback on their use of targeted skills.

Cognitive level of participants. All of the articles reviewed described interventions to treat individuals with ASD without intellectual disability. Most studies required a full-scale IQ of 70 or greater for inclusion in the intervention. Only one study reported a slightly lower threshold, with IQs of greater than 50 required, but all participants were still described as "high-functioning" (Mesibov, 1984). Herbrect et al. (2009) examined the relationship between a participant's IQ and response to a group social skills intervention and unsurprisingly found that higher IQ and language ability predicted better response to treatment. It seems that a specific threshold of intellectual capacity is necessary to optimally benefit from the types of group interventions reviewed, but adaptations for those with more limited intellectual functioning should be considered and tested in the future.

Theoretical basis. While many social skills group were guided by applied behavior analysis (ABA) principles—reinforcement, shaping, and scaffolding of skills (e.g. Dotson et al., 2010; Stichter, Herzog, O'Connor, & Schmidt, 2012; White, Albano, et al., 2010), and some were explicitly behaviorally-based (Mitchell et al., 2010), many group interventions were also informed by other theories. For example, several existing groups were self-described as using cognitive behavioral principles (e.g. White et al., 2013), while others were based on theory of mind (Ozonoff & Miller, 1995) or psychodynamic perspectives (Tyminski & Moore, 2008). However, the majority of social skills group interventions are based in behavioral and cognitive-behavioral principles.

Delivery formats. The format of intervention delivery varied significantly among identified studies. Some studies, such as Dotson et al.'s (2010) teaching interaction procedure, adhered to a set didactic lesson plan, whereas others, such as the group work intervention used by MacKay et al. (2007), use a more activity-based approach. Many groups use a combination of these strategies. For example, Stichter et al. (2010) developed and implemented the Social Competence Intervention (SCI), which addressed social skills through didactics, discussion, modeling, and both structured and naturalistic practice opportunities.

Other teaching strategies used in various groups include modeling (e.g. Laugeson et al., 2012), role plays (e.g. Tse, Strulovitch, Tagalakis, Meng, & Fombonne, 2007), rehearsal opportunities (e.g. Minihan et al., 2011), group discussion (e.g. MacKay et al., 2007), video-feedback (Ozonoff & Miller, 1995), Social Stories (Broderick, Caswell, Gregory, Marzolini, & Wilson, 2002), incentive systems (Mitchell et al., 2010), review of videotaped conversations (Fullerton & Coyne, 1999), visually presented information (Fullerton & Coyne, 1999), interpersonal therapy (Mishna & Muskat, 1998), and even art therapy (Epp, 2008) and music therapy (Hillier, Greher, Poto, & Dougherty, 2012). Of this multitude of teaching strategies, role-play activities were used most frequently and were identified by one researcher as the "most important technique" employed within the intervention paradigm (Mesibov, 1984, p.400).

Only one study compared two different delivery formats: one intended to address social knowledge deficits (Skillstreaming) and the other intended to address social performance deficits (Sociodramatic Affective Relational Intervention; Lerner & Mikami, 2012). The authors found that both interventions resulted in improvements in clinician-rated

socials skills and reciprocated friendship nominations, but in different ways. The results indicated that participants in the social performance intervention rapidly established friendships within the group, whereas participants in the social knowledge intervention gradually formed these friendships. However, at the conclusion of this brief (4 week) intervention, both groups showed equivalent gains in social skills and friendships, suggesting that interventions targeting social knowledge or social performance may have comparable effects on ultimate social competence. Many researchers appear to believe that both methods are useful, as many social skills groups incorporate both didactic and practice opportunities.

Intervention duration and intensity. Research has not reached consensus on the amount and intensity of treatment necessary to obtain clinically significant skill gains, as multiple factors affect intervention efficacy. However, almost half of the investigations identified by this review (46%) described groups that held weekly sessions for approximately 10 – 16 weeks, with each session lasting approximately 40 minutes to 2 hours (e.g. Laugeson et al., 2012; McMahon, Vismara, & Solomon, 2013; Stichter et al., 2012).

Investigations examining short-term social skills interventions (<10 weekly sessions) have yielded limited evidence of effectiveness. One study examining two different social skills group interventions found statistically significant results (based on unblinded interventionist ratings on the SSRS-Teacher Survey) after 90 minutes of treatment every week for four weeks (Lerner & Mikami, 2012). However, the researchers acknowledged the inherent limitations of such unblinded measures and noted that the skill growth observed by the treatment providers was not reported by parents on the SSRS-Parent Survey, thus

concluding that interventions of longer duration were likely necessary to produce generalized gains.

Similarly, Rose and Anketell (2009) reported on a pilot social skills intervention that was delivered through one hour weekly sessions over 5 weeks. Preliminary qualitative information about the usefulness of the intervention was gathered using non-standardized parent questionnaires, but no standardized data was collected to support the efficacy of the intervention. These researchers concluded that ongoing research was likely needed to support the use of this brief intervention. Likewise, an 8-week intervention conducted by Barnhill, Tapscott Cook, Tebbenkamp, and Smith Myles (2002) resulted in parent and participant endorsements of new friendships among group members, but no statistically significant results on standardized measures of emotion recognition or paralanguage skills.

Studies with longer-term implementation (>10 weekly sessions) were associated with a greater likelihood of significant treatment gains. For example, a 12 -week intervention adapted from the Skillstreaming curriculum resulted in significant outcomes with moderate effect sizes (Tse et al., 2007). Similarly, the PEERS curriculum has strong evidence of treatment effectiveness after 12-14 sessions (Laugeson et al., 2009; Laugeson et al., 2012). Many other interventions of this length or longer periods have also found similarly positive results (e.g. Minihan et al., 2011; Mitchell et al., 2010; Stichter et al., 2012). The available evidence seems to indicate that several months of weekly group intervention may serve as a minimum threshold to reliably improve participants' levels of social competence.

An intensive social skills curriculum was noted in a summer program implemented by Lopata et al. (2008). Lopata's summer treatment program took place for 6 hours per day, five days per week over 6 weeks, resulting in a total intervention duration of 180 hours,

much higher than any other studies reported. Moderately large effect sizes were reported on both parent and staff ratings (up to $d = 0.59$) after intervention. While these data are very encouraging, there were noted limitations regarding the lack of blind raters or a no-treatment comparison group.

As an alternative to a set number of group sessions, Tyminski and Moore (2008) implemented a group intervention during which students "graduated" at different rates depending on individual progress. The length of intervention ranged from 4 months to 3 years, with an average duration of 14 months. The results of the study indicated that length of intervention had no impact on outcomes, suggesting that the majority of improvement likely occurred early on in treatment. Further assessment of this hypothesis is warranted and studies that examine dosage effects may shed some light on the optimal balance of efficiency and effectiveness in the use of this treatment modality.

Individualization/Customization. While most studies reviewed remained limited to the group curriculum, a large proportion of the included studies (41%) identified individual target skills for participants and incorporated these into treatment. For example, Mitchell et al. (2010) identified three to five individual target skills for each participant based on their parent's responses to the Social Skills Rating System (SSRS; Gresham & Elliott, 1990) with corroboration from direct behavioral observations. These skills were then embedded into the pre-determined curriculum.

As an alternative means of curriculum adaptation, Minihan, Kinsella, and Honan (2011) described an intervention in which the content of group lessons were dictated by parental responses to the Social Responsiveness Scale (SRS; Constantino & Gruber, 2005).

Based on the SRS, a list of the high priority skills for each participant was combined and used to develop the corresponding group curriculum.

White et al. (2010) used both (a) individual treatment plans and (b) customized group modules. Fifteen modules were available to each group, but the thirteen identified as most applicable in a particular intervention cycle were selected to tailor the content of the group sessions to the specific needs of participants. The researchers identified this strategy as a method to adhere to a manual-based intervention while retaining sufficient flexibility to meet the specific needs of each individual.

Lopata et al. (2008) also described the selection of three or four unique target skills for each child, unrelated to the group curriculum. While all participants in this study received feedback on their target skills, those that were randomized to a response-cost condition earned and lost points based on the use of these specific skills. These points were individually redeemed for edible reinforcers and summed across group members to earn larger incentives (e.g. field trips). These investigators found no significant difference between feedback conditions, but found an overall positive effect for the general intervention package. It is conceivable that individual target behaviors are useful during group interventions, but that the way in which participants receive feedback on these target behaviors is less important.

Interestingly, one of the first published articles on group social skills intervention (Mesibov, 1984) employed an individualized component. This intervention consisted of weekly 60 minute group meetings that were preceded by 30 minute individual meetings for each client, allowing for one-on-one teaching to be paired with group practice. A growing number of empirical investigations are incorporating individualization components into their

curriculum, which may indicate that there is a unique therapeutic benefit to customizing procedures when working with a diverse range of participants.

Inclusion of typically developing peers. Peer–mediated interventions have gained popularity in recent years, with some researchers actively incorporating typical peers into adolescent social skills groups. While no reviewed investigations reported the inclusion of typically developing peers prior to 2004, a number of more recent studies have reported use of this technology (starting with Carter et al., 2004). The majority of the groups identified by this review remain populated solely by teenagers with ASD. However, roughly a third (30%) also included typically developing peers. Some, like the Multimodal Anxiety and Social Skills Intervention (MASSI; White et al., 2013) and the consultation model implemented by Minihan, Kinsella, and Honan (2011), have successfully included typically developing peers with promising results. When peers are involved in group interventions, they are typically trained beforehand and then participate in the groups as social models—providing opportunities for social interaction and peer feedback. Occasionally, peers have been incorporated in different ways. For example, Broderick et al. (2002) described a pilot project in which adolescents concurrently attended a social skills group and a local youth club as a supplemental experience. While no peer models were included in the intervention group itself, every client had weekly scheduled interactions with typical peers during their youth club attendance. Preliminary data reflected possible gains in self-esteem among participants with ASD.

Target skills. The identified empirical studies also varied with regard to specific skills targeted by the intervention. Most of these investigations targeted global social competence, but some focus on a specific area, such as social cognition (Stichter et al.,

2012), theory of mind skills (Ozonoff & Miller, 1995), conversational skills (Dotson et al., 2010), emotional expressiveness (Tyminski & Moore, 2008), social and emotional perspective taking (MacKay et al., 2007), understanding non-verbal communication (Barnhill et al., 2002), and self-determination (Fullerton & Coyne, 1999). More than a few studies specifically targeted anxiety symptoms in conjunction with social skills (e.g. White et al., 2013).

Dependent variables of social change. The intervention approaches covered in this review used a wide variety of dependent/outcome measures to assess the extent to which social improvements had indeed occurred. Most studies (73%) used parent report questionnaires as an outcome measure. Parent measures that were most frequently used included the Social Skills Rating System (SSRS) and the Social Responsiveness Scale (SRS). In addition, some studies used adolescent report questionnaires and/or direct assessments, such as the Diagnostic Analysis of Nonverbal Accuracy (DANVA2). Interestingly, one study compared responses between parents and participants and found that while participants consistently rated their social skills/competence higher that their parents rated them (both before and after intervention), there was a significant positive correlation between the parent and participant ratings (MacKay et al., 2007). This suggests that participant self-report scores may be elevated but generally follow the same intervention trends as corresponding parent ratings. In contrast, however, Tse et al. (2007) found that adolescents reported greater improvements following a group intervention than their parents. Due to these conflicting results, it may be beneficial to continue examining social skill endorsement discrepancies that exist between parent and child.

In addition to parent and adolescent questionnaires, some studies examined results from interventionist ratings (which may be prone to biased endorsements) and teacher ratings (which have poor return rates). A few articles reported the use of ratings by blinded clinicians (e.g. Herbrecht et al., 2009; White et al., 2013), which represents one of the most rigorous forms of evaluation, as they are completed by experts unaware of the treatment status of participants.

A few studies examined direct, observable data from participants. McMahon, Vismara, and Solomon (2013) examined behavioral data from the group sessions themselves, documenting increases in the number of responding vocalizations made by group members. Similarly, Mitchell et al. (2010) conducted behavioral probes during training sessions and observed increases in the participants' target skills, which included operationally defined introductions, initiated conversations, problem-solving skills, and group joining skills. As a limitation, most of the studies that examined behavioral data collected it solely during the social skill group sessions. For example, Dotson et al. (2010) measured behavioral change during group sessions and also collected generalization data during these group sessions (i.e. participants participated in brief side conversations with a typical peer who was also involved in the weekly group).

One study was identified that collected behavioral social data outside of the group sessions. Mitchell et al. (2010) gathered generalization data by creating natural opportunities for social probes in alternative locations within the same building, with people who were unrelated to the participant or the study. While this represents progress in the objective behavioral measurement of generalization, there is substantial room for improvement. In the future, researchers may increasingly consider using behaviorally-based

social probes conducted more distally from the site of intervention and importantly, use of interaction partners who are unknown to participants and of a similar age.

Generalization and maintenance techniques. While most of the identified studies primarily focused on immediate treatment effects, more recent investigations have also attended to generalization and maintenance considerations. Assessing whether skills gains are (a) appropriately generalizing to other settings and (b) maintaining beyond the duration of treatment are arguably the most important considerations in any intervention research (Stokes and Baer, 1977). Specific techniques have been incorporated into existing projects to promote generalization and maintenance. For example, many interventions include follow-up homework assignments (e.g. McMahon at al., 2013). While homework assignments vary between interventions, they generally involve in-home or community practice of the skill(s) addressed in the group session. Some studies required comprehensive homework assignments consisting of both (a) skill practice guidelines and (b) written responses to socially relevant questions (Mitchell et al., 2010). In addition, these researchers implemented an incentive system to encourage consistent homework completion. Homework assignments have been described as "key" to obtaining generalization of skills by Frankel et al. (2010), as they encourage application of core socialization techniques outside of the time-limited context of each group session.

Another strategy intended to promote generalization and maintenance of skills is the inclusion of a parent component. Parents who are either directly involved in the treatment or informed about the skills being targeted may be in a better position to assist their child in accurate use of these skills outside of the group, thereby leading to widespread, generalized use. Laugeson et al. (2009; 2012) evaluated the Program for the Education and Enrichment

of Relational Skills (PEERS) program, which included both homework and parent components. Specifically, a concurrent parent group is conducted alongside the adolescent group, in which parents are provided with education and information about the same skill curriculum as their children. These studies found significant changes to social functioning that maintained after treatment. Generalization measures (completed by teachers) were also significant 14 weeks after intervention, providing evidence of both generalization and maintenance of the skills learned during the initial PEERS intervention.

Other researchers have incorporated parents into the treatment in different ways. For example, Mitchell et al. (2010) conducted parent education sessions every three weeks of their 12-week adolescent group intervention. These parent-oriented sessions consisted of psychoeducation, instruction of pragmatic strategies to enhance and generalize their child's target social skills, and review of video recordings from the group sessions to illustrate behavioral strategies to use at home. In another study, parents met with the group leaders three times over the course of the 11-month intervention to exchange information related to intervention progress and current difficulties (Herbrecht et al., 2009). Several articles indicated the need to increase parent participation within the intervention process, with several proposing future inclusion of parent training components (e.g. Corbett et al., 2011; Tse et al., 2007).

In addition to home practice and parent feedback meetings, MacKay et al. (2007) reported using community outings as a generalization technique. These researchers suggested that outings provided participants with opportunities to practice their skills in a naturalistic setting, and they found statistically significant results on all outcomes measures.

Across multiple investigations, there is a growing consensus that the ultimate goal of treatment must focus on generalization considerations.

Summary. There is a rapidly growing evidence base for group interventions that promote social skills in adolescents with ASD. While this population has historically received less research attention, the field has rapidly expanded in response to the critical socialization needs of these individuals.

Findings from this review indicate several key themes and trends across social skills groups. Most of the articles identified in this review evaluated program effectiveness using a pre-post design. However, researchers appear to be transitioning from these preliminary feasibility studies to more rigorously controlled trials, with several RCTs having been conducted and published in recent years (e.g. Laugeson et al., 2012; Lerner & Mikami, 2012; White et al., 2013). With regard to treatment setting, most of the studies examined in this review took place in clinic settings, but a small number of studies suggest that social skills groups may be effective in school and community settings (e.g. Barnhill et al., 2002; Kempe & Tissot, 2012). The majority of studies were conducted in the United States, but with growing number of articles describing international participants and projects (e.g. Herbrecht et al., 2009), which provide preliminary evidence for the generalizability of this form of intervention to a variety of cultures and contexts.

Presently, no studies were identified that assessed the relative merit of smaller versus larger groups of participants. However, smaller groups of 4 to 6 participants (e.g. Epp, 2008; Stichter et al., 2010; White et al., 2013) appear to be the current preference, likely because they allow the researchers a logically manageable group, with the flexibility to conduct both experiential and practice-based techniques. There is, however, evidence that

larger groups may be feasible, particularly when delivering didactic lessons (e.g. Laugeson et al. 2012). In terms of participant characteristics, this review did not identify any studies that evaluated social skills group efficacy for individuals with co-occurring intellectual disability, and the findings from several studies seem indicative of certain cognitive pre-requisites to adequately benefit from the described programs. However, adaptations for individuals with more limited intellectual capacities will undoubtedly need empirical examination in the future.

The majority of groups reviewed were theoretically based in broad behavioral or cognitive-behavioral principles (e.g. White et al., 2013), although a number of group interventions were based on theory of mind and other theoretical orientations (e.g. Begeer et al., 2010; Ozonoff & Miller, 1995). While the various delivery formats implemented in each study initially appear to be very unique, most interventions include some combination of didactic teaching and experiential components (e.g. group discussion and/or rehearsal opportunities). Role-plays were specifically identified as a common and useful treatment technique. In terms of social competencies, most of the reviewed interventions targeted social skills comprehensively, while a smaller number of studies identified a more specific aspect of social competence, such as social cognition or non-verbal communication (e.g. Barnhill et al., 2002). A number of studies targeted anxiety symptoms in addition to social skills (e.g. White et al., 2010; White et al., 2013), suggesting that the performance deficits associated with co-occurring anxiety symptoms are a core intervention consideration. Mood symptoms, and specifically how these traits interact with social performance, must be considered when developing future interventions for adolescents on the spectrum.

Currently, social skills program treatment intensity ranges from 6 hours up to 180 hours of total intervention, with the majority of programs taking place weekly for 10 – 16 weeks. The available evidence appears to indicate that several months of weekly group intervention is the minimum amount necessary to reliably improve participants' social skills. Treatments of shorter duration yielded more limited evidence of social gains and generalization of these skills (e.g. Barnhill et al. 2002; Rose & Anketell, 2009), while treatments of a significantly longer duration do not yet use proper assessment methods and/or present compelling data to support prolonged participation in these programs (e.g. Tyminski & Moore, 2008). A high priority for researchers should be to empirically examine the benefits of continued participation in groups beyond the typical 10 – 16 week range to determine whether ongoing participation leads to measurable social competence gains. In addition, ensuring that the maximum amount of benefit is balanced with a high level of efficiency will be important, particularly given the current health care climate, which clearly calls for time-limited and low-cost treatment approaches. For this reason, studies that examine dosage effects are in particular demand in order to guide our understanding of how much or how little treatment can provide that optimal balance between efficacy and efficiency.

A trend toward providing a degree of programmatic flexibility and customization within the group curriculum was noted among many of the reviewed investigations. Individualization was achieved in a variety of ways, including treatment plans or target behaviors unique to each participant or customization of the broader group topics based on individual participant needs (e.g. Lopata et al., 2008; Minihan et al., 2010; Rose & Anketell, 2009). These modifications allow researchers and clinicians to a) target a large number of

participants through the use of a group intervention, and also b) address the areas of greatest need for each individual. This blend of group and individual treatment optimizes the efficient delivery of services, and may become a strategy used more and more widely in the current era of managed health care restrictions.

The use of generalization and maintenance techniques also differed between interventions. Some common techniques included homework assignments and varying degrees of parental and peer involvement. Of particular importance, an emerging trend was noted toward the incorporation of typically developing peers into the group interventions in recent years (e.g. Corbett et al., 2011; Dotson et al., 2010; Mackay et al., 2007). Peer-mediated treatment strategies appear to be growing in both popularity and associated research evidence, with this component increasingly changing the landscape of social skills groups. Positive findings from these investigations indicate that researchers may wish to consider including typically developing peers as models and facilitators in group interventions, particularly given the growing evidence base for peer mediated interventions (Watkins et al., 2015; Zhang & Wheeler, 2011).

Presently, the literature suggests that social skills group interventions can be effective. Efficacy has been measured through a variety of means, most commonly through parent report measures (e.g. Gabriels et al., 2012; Shin et al., 2011; Valenti et al., 2010). Very few studies reported the use of rigorous measurement techniques such as blinded clinician ratings or objectively coded naturalistic behavioral observations (e.g. Koegel et al., 2012; McMahon et al., 2013). As the initial efficacy results indicate the usefulness of social skills groups for adolescents with ASD, these more advanced measurements of change will be important to include in the next wave of intervention studies.

The START Group

The literature review described above points to a number of areas of weakness in existing groups that can be improved upon. The START group was developed specifically to address some of the noted areas for improvement. Several areas were identified that were felt to be key to maximizing treatment effects, including the use of an experiential format, the inclusion of typically developing peers, the incorporation of self-management and motivational techniques, the use of individualized target skills, the inclusion of a parent-education component, and weekly homework assignments to promote generalization and maintenance. In addition, when developing the START group, Krasny, Williams, Provencal, and Ozonoff's (2003) recommendations for a model social skills group curriculum were followed. They identified nine "essential ingredients" which have been incorporated into the START program:

1. Make the abstract concrete,
2. Provide structure and predictability,
3. Provide scaffolded language support,
4. Provide multiple and varied learning opportunities,
5. Include "other"-focused activities,
6. Foster self-awareness and self-esteem,
7. Select relevant goals,
8. Program in a sequential and progressive manner, and
9. Provide opportunities for programmed generalization and ongoing practice. (p. 111)

Effectiveness of the START group has previously been examined through a small within-subjects pilot study. This study examined outcomes (using a number of different measures) after a 5 week waitlist and after participation in the START group in 5 week increments, up to 20 weeks of treatment. While the group was far too small to determine statistical significance, visual inspection of data graphs showed no improvement during the waitlist and clear improvement in all measures during treatment. Effect sizes (Cohen's *d*) for parent report measures ranged from 0.46 (medium) to 1.61 (large). For participant report measures they ranged from 0.20 (small) to 0.74 (large). Behavioral measures also showed improvement with effect sizes ranging from 0.41 (medium) to 3.30 (large) depending on participant and behavior measured. This data gives an initial idea of the usefulness of the START group and has led to the development of a larger randomized controlled trial (RCT) to examine efficacy in a larger sample.

The Current Study

Results from the literature review not only informed the development of the START group, they also influenced the design of the current study. Specifically, the use of an RCT, the use of rigorous behavioral and blinded outcome measures, and the examination of dosage effects have all been employed. The RCT has collected data on 19 participants and compares a treatment group to a waitlist control group. While the current study will examine the data from these 19 participants, positive results from this preliminary investigation will justify the continuation of the RCT to eventually collect data on approximately 40 participants. While the initial aim of this research study is to examine the effectiveness of the START group by comparing the treatment group to the waitlist control, additional research questions have also been developed. Of particular interest to this

researcher is the effect of dosage, or length of treatment, on the social skill improvement of participants. An investigation of this type was not found in the literature and the importance of length of treatment remains to be studied. This can be examined by analyzing any differences found between outcome measures administered during 5-week increments throughout the intervention. A number of outcome measures are being collected, including parent-report measures, participant report measures, behavioral observations, and subjective ratings by naïve raters. The proposed research study therefore seeks to answer the following research questions:

1. Are the social skills of the participants in the START treatment group significantly greater than those of the participants in the waitlist control group after completion of the 20 week program?
    a. According to parent report measures of social skills
    b. According to participant self-report measures of social skills
    c. According to objective behavioral observations of social skills
    d. According to naïve raters' subjective ratings of social skills
2. At what point during the 20 week treatment period does a significant gain first occur?
    a. According to parent report measures of social skills
    b. According to participant self-report measures of social skills
    c. According to objective behavioral observations of social skills
    d. According to naïve raters' subjective ratings of social skills

3. After significant gains are first observed, do significant gains continue to be made, or at what point during the 20 week treatment period are significant gains no longer achieved?
    a. According to parent report measures of social skills
    b. According to participant self-report measures of social skills
    c. According to objective behavioral observations of social skills
    d. According to naïve raters' subjective ratings of social skills

Method

Participants

Nineteen adolescents, aged 12 to 18, with a previous diagnosis of an Autism Spectrum Disorder (Autistic Disorder, Asperger's Disorder, or Pervasive Developmental Disorder, NOS) participated in this study. Additional inclusion criteria included a Verbal IQ greater than 70, the use of full sentence communication, and a Social Responsiveness Scale (SRS) score greater than or equal to 60 (indicating the presence of an ASD).

Participants were recruited from several sources, including the identification of current and past clients of the Koegel Autism Center who meet inclusion criteria, referrals from community organizations and schools, and responses to flyers, email announcements, advertisements, and other information dissemination efforts.

Ten subjects were male and nine were female. The average age of participants at baseline was 13.8 years (ranging from 12 to 17 years). In addition to an ASD diagnosis, five participants had existing co-morbid diagnoses, including Anxiety, Social Anxiety, Attention Deficit Hyperactivity Disorder, and Oppositional Defiant Disorder. Twelve participants

identified their ethnicity as Caucasian, four as Hispanic/Latino(a), one as Asian, one as Middle Eastern, and one as mixed ethnicity. Thirteen were enrolled in a public school setting, three in a private school setting, and three were homeschooled. Of those not homeschooled, 10 were mainstreamed, two were in special education classes, and four were in a combination of regular and special education classes.

Three participants dropped out of the study prior to completion. One dropped before treatment began, one dropped after missing four groups due to an extra-curricular scheduling conflict, and one dropped after attending two groups due to both a scheduling conflict and a disinterest in continuing.

Research Design

The current study employed a two-part research design. First, a randomized, waitlist controlled, between-group design was used to examine the effectiveness of the START group intervention to answer research question #1. Based on the results of this initial investigation of efficacy, research questions #2 and #3 were examined by employing a within-subjects repeated measures design in order to understand the social improvement trajectories of participants as they progressed through the experimental intervention.

Measures

Dependent measures included parent-report survey measures, participant-report survey measures, objective behavioral observations of conversation samples, and subjective ratings of conversation samples as complementary measures of social skill improvements. In addition, an IQ test was administered to determine eligibility for participation in the study.

Kaufman Brief Intelligence Test, Second Edition. The KBIT-2 was chosen as a measure of intelligence due to the brief nature of the administration. While many IQ measures take hours to administer, the KBIT-2 takes only 20 minutes to administer. Detailed information about each participant's IQ was not necessary for the purposes of the study, but the KBIT-2 allowed us to confirm that the participant's cognitive ability would permit them to fully participate in the intervention. The KBIT-2 manual reports internal consistency coefficients ranging from 0.89 to 0.96 and averaging 0.93. Test-retest reliability ranges from 0.88 to 0.92. In addition, the KBIT was shown to correlate well with other measures of cognitive ability, including the WASI, WISC-IV, WAIS-III, and KABC-2.

Social Motivation & Competencies Scale (SMCS). The SMCS is an unpublished rating scale developed by the researchers for use in this study as the primary outcome measure. Separate forms were designed for both participants and their parents. Items pertaining to social comfort, conversation skill use, empathy, friendships, appropriate behavior, social contact, and social interest are rated on a 1-5 Likert-type scale and a total score is obtained.

Social Responsiveness Scale, Second Edition (SRS-2). The SRS-2 is a rating scale which measures the severity of autism spectrum symptoms as they occur in natural social settings. It was completed by a parent and used as an outcome measure in the current study. Internal consistency was reported at 0.95 and test-retest reliability was reported at 0.88 to 0.98. Convergent validity was also established with similar scales such as the SCQ, CCC, and SCDC. The SRS provides a total score as well as a Social Communication and Interaction subscale and a Restricted Interests and Repetitive Behaviors subscale.

Social Skills Improvement System Rating Scales (SSIS). The SSIS is a rating scale which measures several aspects of social skills, including Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, and Self-Control. Separate rating scales are available for clients, parents, and teachers. Both the parent and client forms were used as outcome measures in the current study. Coefficient alphas ranged from 0.95 to 0.96 for the parent form and fell at 0.95 for the client form. Test-retest reliability was reported at 0.86 for the parent form and 0.80 for the client form. Modest support for both convergent and discriminant validity has also been reported. In addition to an overall score, the SSIS provides a number of subscales, including Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, Self-Control, Externalizing, Bullying, Hyperactivity/Inattention, Internalizing, and Autism Spectrum scales.

Objective behavioral observations of conversation samples. At each data collection time point, two brief (5 minute) conversations probes between the participant and two unfamiliar, similar aged peers (separately) were video-recorded. These recordings were then objectively coded by raters naïve to the hypothesis of the study. To control for observer drift, videotapes were scored in random order. Specific behavioral measures were operationally defined and reliability between coders was established. The two behaviors measured in the current study were percentage of questions asked and mutual engagement between the participant and their conversation partner. Questions were defined as a verbal query that expects a response from the conversational partner. Mutual engagement was defined as a verbal utterance of more than two words coded for both the participant and the conversation partner within the same 5-second interval. Interrater reliability was calculated for 28% of the videos and was calculated using the standard formula for percent agreement:

observer agreements divided by the total number of agreements plus disagreements multiplied by 100%. Reliability percentages averaged 88% (range 80–100%) on the number of questions asked. Reliability on mutual engagement was also calculated at 88% (range 80–100%).

Subjective ratings of conversation samples. Additionally, subjective ratings were completed by other naïve raters. These raters were asked to rate each person in the videos (i.e. both the client and the peer) on a 1 – 10 Likert-type scale on four subjective factors, including how strong the person's social skills appeared to be, how comfortable the person seemed to be conversing, how socially awkward the person seemed, and how likely the person was to have many friends. These raters received no training and were not given operationalized definitions, as the intention of these ratings was to determine the social significance of the intervention.

It was expected that due to the subjective nature of these ratings, there was likely to be a high degree of inter-rater variability, which was likely to result in variable scores impacted heavily by the fact that different raters were being used at different time points for the same participant. Because of this acknowledged source of error, an alternative method of subjective rating (paired subjective ratings) was also employed. In this method, a rater was asked to rate two videos of a client and was not asked to rate the conversational peer. This method, therefore, reduced interrater variability, as the same rater provided a score for both the Baseline and Week 20 time points.

Procedure

Families who contacted us with interest in participating in this social skills group intervention were scheduled for an individual intake appointment to determine eligibility

based on our inclusion criteria. Adolescents meeting criteria were then randomly assigned to either the treatment or waitlist control condition, with randomization blocked by gender to ensure equal distribution of gender across both groups. The treatment group was enrolled in the group intervention immediately following the baseline assessment and the waitlist control group was enrolled in the group intervention following a 20 week waiting period and a second baseline assessment.

Baseline/intake appointment. All participants were asked to come in for an initial session which consisted of obtaining consent/assent from parents and participants and collection of baseline measures including demographics, IQ, SRS, SSIS, SMCS and video-recorded conversations (one with a male peer and another with a female peer).

Waitlist control. All participants assigned to the waitlist control group (and their parents) were asked to complete the outcome measures again (SRS, SSIS, SMCS, and video-recorded conversations) after a 20 week waiting period. They then received the intervention (described below) to ensure that all participants were given access to treatment.

Treatment group. All participants assigned to the treatment group began the intervention immediately after completion of their baseline measures. An individual target skill was identified for each participant, based on a rank ordering of social difficulties completed separately by the parent, participant, and intake clinician. This target skill was discussed with the participant and served as the focus of the participant's self-management during group sessions. Examples of targeted skills include asking on-topic questions, making positive statements, elaborating on statements, initiating conversations, and respectfully disagreeing. These goals were reevaluated every five weeks during progress

meetings and new skills were targeted as participants demonstrated mastery of the previous target skill.

Groups occurred once per week and lasted approximately 75 minutes. They consisted of the following components: an individual check-in session, unstructured socialization with self-management, a group activity, a group discussion of a social skill topic, and an individual check-out session with parent involvement.

The check-in session primarily consisted of a review of the participant's homework goals from the previous week. Achievement of their goals (as well as attempts) were praised by the clinician. Discussion of any difficulties and/or successes also took place. Lastly, the check-in sessions were used to prime the participant for the group activities through a short description of the week's topic and activity and a brief practice conversation with the clinician while the participant self-managed for their target skill.

After completing the individual check-in sessions, the participants were brought together as a group with the clinicians and the typically-developing high school peers. The first 20 minutes of the group consisted of unstructured socialization. Participants and peers conversed with each other while self-managing for their individual target skills. Clinicians praised participants for use of their skill as well as their use of self-management. This unstructured time was intended to help increase the participants' comfort in the group as well as to provide an opportunity to practice their target skills in a naturalistic context.

After approximately 20 minutes, the group transitioned into an activity. Activities varied each week, but were intended to foster the sharing of personal information, learning the interests of other participants, building comfort, and promoting cooperation and

teamwork. Activities were also intended to be enjoyable in order to assist with retention of participants.

Once the activity was completed, the clinicians began a brief discussion of the previous week's topic to facilitate retention of the material. Participants were asked to discuss any difficulties or successes they experienced when practicing the skill during their homework. This was followed by a facilitated discussion of the current week's social skill topic. The topic was introduced and the participants were asked to provide relevant suggestions on how to successfully use a particular social skill. All suggestions were recorded on a large pad of paper to serve as a visual aide. Clinicians used role plays and video clips to demonstrate both good and bad examples of social behavior. Participants were asked to comment on aspects of the role players' or characters' behavior that was positive and aspects that were not successful, while giving suggestions for ways to improve their behavior. After further discussion of the topic, participants and peers were asked to practice using the discussed skill. The practice component was conducted in a number of different ways that varied each week and included pairing up with a neighbor to practice, rotating through various partners for practice and feedback, or practice through role-playing a scenario. This portion of the group was intended to promote understanding of a social skill topic, as well as to provide opportunities to both observe and practice the skill. A list of the weekly topics can be found below in Appendix A.

Lastly, individual check-out sessions were conducted with each participant and their parent(s). Participants were asked to describe the topic of the group to their parent(s) in order to support their own understanding of the topic as well as to allow their parent(s) to have an understanding of the content of the group (with the added benefit of increasing the

likelihood of their parents promoting this skill at home after hearing about our discussion of it). Parents were also provided with a written description of the topic to take home. Lastly, the clinician, participant, and parent(s) jointly established homework goals for the following week, which were based upon the participant's individual target skill as well as the topic of discussion that week.

Adherence to this treatment protocol was monitored through the use of fidelity checklists. Fidelity to the intervention was assessed by one of the group leaders each week, using the checklist to record all completed components of the intervention as they took place during the group.

Data collection. Outcome data was collected during individual sessions conducted once every five weeks (SMCS, SRS, SSIS, and video-recorded conversations).

Results

Baseline Characteristics

Table 1 presents the mean demographic and baseline variables for each group. Chi square analyses for percent male, percent enrolled in public school, percent mainstreamed, and percent Caucasian were not significant. *T*-tests for age, grade, ongoing number of weekly intervention hours, KBIT-2 Verbal IQ score, and outcome variable baseline scores all failed to reach significance, with one exception: the participant version of the SSIS was significantly higher in the waitlist group ($p = 0.02$). However, it should be noted that the waitlist group did appear to have higher baseline scores on all of the parent report and participant self-report outcome variables, although these group differences were for the most part non-significant. While these results are interesting to note, any between group differences will be statistically accounted for by the use of an ANCOVA for further analyses.

Table 1. *Mean Demographic and Baseline Variables for Treatment and Waitlist Control Groups*

	Treatment Group (n = 9)	**Waitlist Control Group** (n = 8)	***p*-value**
Age	13.7	14.1	ns
Grade	8.1	8.8	ns
Percent Male	44	63	ns
Number of Weekly Intervention Hours	2.6	2.9	ns
Percent in Public School	56	88	ns
Percent Mainstreamed	67	63	ns
Percent Caucasian	56	63	ns
KBIT-2 Verbal IQ	97.4	90.9	ns
Parent Surveys - Baseline			
SMCS	72.3	75.9	ns
SSIS	71.7	81.8	ns
SRS	76.0	76.1	ns
Participant Surveys - Baseline			
SMCS	95.6	98.6	ns
SSIS	84.8	108.4	0.02
Behavioral Observations - Baseline			
Percentage of Questions Asked	21.5	13.6	ns
Percentage of Speaking	43.6	44.7	ns
Percentage of Mutual Engagement	14.8	12.9	ns
Subjective Ratings - Baseline	21.5	17.5	ns

Fidelity of Implementation

Fidelity to the treatment protocol was operationalized as the percentage of intended treatment components correctly delivered within the session. Fidelity was measured in 83% of the weekly sessions. Fidelity ranged each week from 80-100%, with an average of 98% of treatment components delivered each week, indicating a high degree of adherence to the protocol.

Between-Subjects ANCOVA

To answer research question #1, whether the scores of the participants in the START treatment group were significantly greater than those of the participants in the waitlist control group after completion of the 20 week program, a between-subjects analysis using ANCOVA was performed. The ANCOVA was conducted to analyze post-test differences (week 20) while controlling for pre-test scores (baseline).

Data screening. Before performing any analyses, the data were screened to ensure that all assumptions of the ANCOVA were met. The assumptions of an ANCOVA include independence of observations, normality of distribution, homogeneity of variance, reliability of covariates, linearity, and homogeneity of regression. The independence of observations assumption is met through the randomization inherent in the research design. The reliability of covariates assumption is met as much as possible through the use of reliable baseline measures. Both the SSIS and SRS have high reported reliability and interrater reliability was established for all behavioral observations.

Normality of distribution was assessed through the examination of histograms and boxplots. Histograms were obtained for each of the outcome measures (Week 20 scores) as well as each of the covariates (Baseline scores). Additionally, histograms were obtained for

the entire sample as well as for the treatment and waitlist groups separately. Boxplots were also obtained for each of the outcome measures and covariates with results displayed separately for the treatment and waitlist groups. All histograms indicated approximately normal distributions, although the data from the behavioral observations of question asking indicated that both the outcome and covariate data for this measure were slightly positively skewed. However, ANCOVAs are known to be fairly robust to moderate deviations from normality, particularly when the sample sizes are not unbalanced, which is the case in this design (Glass, Peckham, & Sanders, 1972). While the majority of boxplots did not indicate any outliers, the following outliers were identified. One outlier was found in the Parent SMCS covariate boxplot (of baseline scores) from the treatment group. A different case was identified as an outlier in the data from the covariate (baseline) scores on the SRS, also from the treatment group. The Participant version of the SMCS also contained a single outlier in the dependent variable (Week 20 scores) from the treatment group. The Participant version of the SSIS contained three outliers, one in the outcome data from the waitlist group, and two from the covariate data in the treatment group. The question asking data from the behavioral observations also contained a single outlier in the covariate data from the treatment group. The behavioral observations for the percentage of time mutually engaged contained three outliers in the covariate data from the treatment group. Because of the moderately large number of outlying scores, the subsequent analyses were first performed including the outlying scores, with follow-up analyses conducted to assess for any noteworthy changes that occurred when excluding the relevant outliers.

The ANCOVA model also assumes linearity of the relationship between the covariate and the dependent variables. This assumption was assessed through examining

bivariate scatterplots, all of which indicated approximately linear relationships. In one case, on the Participant version of the SSIS, two cases appeared to fall outside of the otherwise linear trend. These two cases corresponded with outlier cases identified earlier using boxplots, which were then removed from the secondary analysis.

Homogeneity of variance, another ANCOVA assumption, was assessed using Levene's test. In all cases, the Levene statistic was not significant, meaning that this assumption was not violated for any of the outcome measures. Lastly, the ANCOVA model assumes homogeneity of regression, which was assessed through evaluating the significance of the interaction term between the baseline score (covariate) and the treatment group. Again, this term was non-significant in all cases and the assumption was not violated.

Between-subjects analysis with outliers included. Each of the primary and secondary measures were not found to be statistically significant when controlling for baseline scores. See Table 2 for the *p*-values obtained from each test as well as the effect sizes, which were based on the partial eta squared obtained from the ANCOVA and are reported as Cohen's *d* (Cohen, 1992). However, despite the statistical non-significance of these results, clinical significance is indicated in many cases by notable effect sizes.

Table 2. *ANCOVA Results*

	***p*-value**	**Cohen's *d***
Parent Surveys		
SMCS	0.23	0.67
SSIS	0.99	0.00
SRS	0.23	0.68
Participant Surveys		
SMCS	0.17	0.77
SSIS	0.75	0.17
Behavioral Observations		
Percentage of Questions Asked	0.27	0.61
Percentage of Mutual Engagement	0.60	0.28
Subjective Ratings	0.88	0.09

Between-subjects analysis with outliers removed. Based on the observed outlying scores in most of the outcome measure data, all ANCOVA analyses were repeated with outliers removed. Again, all of the results were statistically non-significant. However, a noted improvement in *p*- values and the associated effect sizes was observed in five of the eight outcome measures, as seen in Table 3. A majority of the resulting effect sizes fell in the clinically significant range, indicating a promising treatment effect. As no outliers were found in the data for subjective ratings, no new analyses were performed for this measure.

Table 3. *ANCOVA Results with Outliers Removed*

	***p*-value**	**Cohen's *d***
Parent Surveys		
SMCS	0.14[a]	0.88[a]
SSIS	0.89[a]	0.09[a]
SRS	0.19[a]	0.77[a]
Participant Surveys		
SMCS	0.08[a]	1.07[a]
SSIS	0.42[a]	0.51[a]
Behavioral Observations		
Percentage of Questions Asked	0.33	0.56
Percentage of Mutual Engagement	0.63	0.29
Subjective Ratings	na	na

a. Indicates a statistical improvement when outliers were removed from the data

Post-hoc power analyses. Given the lack of statistical significance in all outcome measures, post-hoc power analyses were conducted using G*Power 3.1.5 in order to determine the observed power for each test. The alpha level was set at 0.05 and effect sizes were based on the partial eta squared term obtained from each ANCOVA. The sample size was set at 17, the number of groups was set at 2 (waitlist and treatment), the numerator degrees of freedom was set at 1 (to test for a main effect of group; number of levels minus one) and the number of covariates was set at 1 (the baseline scores for each outcome measure). The calculation resulted in an observed power for some outcome measures as low as 0.05, with the highest power observed for the SMCS Participant measure at 0.54 (see Table 4).

Table 4. *Observed Power of ANCOVA Analyses*

	with outliers	outliers removed
Parent Surveys		
SMCS	0.25	0.40
SSIS	0.05	0.05
SRS	0.26	0.32
Participant Surveys		
SMCS	0.39	0.54
SSIS	0.06	0.16
Behavioral Observations		
Percentage of Questions Asked	0.22	0.19
Percentage of Mutual Engagement	0.09	0.09
Subjective Ratings	0.05	na

With power ideally falling above 0.8, it is clear from these calculations that the between-subjects portion of this study was significantly underpowered. In most cases, a slight improvement in power was noted for the analyses performed with outliers removed, but all tests remained underpowered. Therefore, rather than basing conclusions about the above ANCOVA results on *p*-values alone, effect size was calculated to provide further information about the impact of the intervention.

Effect sizes. The effect size of each outcome variable, as seen in Tables 2 and 3, was calculated using the partial eta squared results from the ANCOVA and converting to Cohen's *d*. Because of the particularly low levels of power associated with the analyses above and the improved power in the analyses performed with outliers removed, the effect sizes observed when outliers were removed will be discussed here. The primary parent report measure, the SMCS Parent, was found to have a large effect size ($d = 0.88$).

Similarly, the primary participant report measure (the SMCS Participant) was also found to have a large effect size (d = 1.07). Most other measures exhibited small to medium effect sizes ranging from 0.29 to 0.77. These results indicate a clinically significant treatment effect after participation in the START group. This positive effect is captured by a variety of outcome measures, including parent report, participant self-report, and objective behavioral observations.

Two measures, the SSIS Parent and the subjective ratings, indicated negligible effects (both d = 0.09). Based on these effect sizes, further investigation to address research questions #2 and #3 using a repeated measures design is supported for some, but not all, of the outcome measures. A cutoff of d = 0.2 was employed to determine whether further investigation was warranted. While this cutoff is somewhat arbitrary, it represents Cohen's cutoff for a "small" effect size and in the current study was also felt to adequately separate those measures that appeared to capture some promising effects (ranging from small to large) from those measures that appeared to display very minimal or no change during treatment. Based on this cutoff, six of the eight original outcome measures were analyzed to assess for treatment trajectories. Because the SSIS Parent and the subjective ratings did not appear to be sensitive to the changes observed in the other measures, these two measures were not included in the subsequent analyses.

***T*-test for subjective rating alternative procedure.** As described in the procedures, the method for obtaining the subjective rating data inherently created a large degree of error, as different raters were employed to rate each time point for a given participant. Because of rater's large variability in their subjective impressions, this creates some difficulty in interpreting the results. To address this problem, an alternative method was employed in

which a rater was asked to rate the same participant both before and after treatment (although the rater was naïve to the treatment status of the participant, as well as to the intervention, knowing only that they were rating the same individual at two different time points). This method, therefore, controlled for the error associated with the original method. A *t*-test was completed comparing the pre- and post-intervention ratings and was found to be statistically significant ($p = 0.001$), with an effect size of $d = 0.41$. These results indicate that individuals unfamiliar with ASD were able to observe noticeable changes in participants' conversational skills and that this paired rating procedure may be a more acceptable way to collect subjective ratings. The medium effect size and strong statistical significance associated with this new outcome measure provides additional support for the efficacy of the intervention, as well as the social significance of the observed improvements.

Within-Subjects Repeated Measures ANOVA

To answer research questions #2 and #3, (at what point during the 20 week treatment period does a significant gain first occur? And after significant gains are first observed, do significant gains continue to be made?), a within-subjects analysis using a repeated measures ANOVA (rANOVA) was performed. This analysis was only performed for outcome measures that had a minimum effect size of $d = 0.2$ based on the between-subjects ANCOVA results. For this reason, the SSIS Parent questionnaire and the subjective ratings were excluded from this analysis.

Power analysis. The a priori power calculation for the within-subjects data, setting the number of groups at 1, number of measurements at 5, and correlation among repeated measures at 0.8, resulted in a required sample size of only 4 participants in order to have

adequate power to detect significant results. The study's sample size of 16 should therefore have more than adequate power to detect significant results using an rANOVA analysis.

Data screening. Assumptions of rANOVAs include the assumptions previously tested for an ANCOVA, in addition to the assumption of sphericity. Sphericity was evaluated using Mauchley's Test of Sphericity. For three of the six analyses run, the assumption was met, as indicated by a non-significant *p*-value on Mauchley's test. In three cases, the assumption of sphericity was violated (for the SMCS Parent measure, the SSIS Participant measure, and the behavioral observation of mutual engagement). When the assumption of sphericity is not met, two options exist for analysis of the data. If the epsilon term (a measure of the degree to which the data violate sphericity) is less than 0.75, a Greenhouse-Geisser adjustment is often used (Greenhouse & Geisser, 1959), and if epsilon is greater than 0.75, a Huynh-Feldt adjustment is often used (Huynh & Feldt, 1976). Both of these techniques make downward adjustments to the degrees of freedom, which creates a more conservative test of significance and decreases the chance of a Type I error, however, the Greenhouse-Geisser adjustment is more conservative than the Huynh-Feldt adjustment (and therefore used in cases of greater violations of sphericity) (Warner, 2008). In the case of the current study, two of the sphericity violations resulted in epsilon terms below 0.75, which required Greenhouse-Geisser adjustments (SMCS Parent and SSIS Participant) and one violation resulted in an epsilon term greater than 0.75 in which case a Huynh-Feldt adjustment was applied (mutual engagement).

It was additionally necessary to re-evaluate normality, as data from additional time points were to be included in this analysis. The data were again found to largely fall within normal limits, with only minor positive skews in the question asking data. Outliers were

noted in the following measures: SMCS Parent (three outliers at Week 10), SMCS Participant (one outlier at Week 20), SSIS participant (one outlier at baseline), question asking (two outliers: one at baseline and another at baseline, Week 5, and Week 20), and mutual engagement (two outliers at baseline). As was done in the previous analysis, the following rANOVA results will first be presented with all outlying scores included, with follow-up analyses presented to assess for important changes when outliers were removed from the analysis.

Repeated measures analysis with outliers included. Five of the six rANOVA analyses resulted in statistically significant results, as seen in Table 5. Each of these statistically significant results was associated with a very large effect size.

Table 5. *rANOVA Results*

	***p*-value**	**Cohen's *d***
Parent Surveys		
SMCS	<0.001	1.62
SRS	<0.001	1.33
Participant Surveys		
SMCS	0.004	1.08
SSIS	0.038	0.96
Behavioral Observations		
Percentage of Questions Asked	0.231	0.59
Percentage of Mutual Engagement	0.001	1.25

Repeated measures analysis with outliers removed. When outliers were removed from the analysis, an improvement was observed in many of the *p*-values and effect sizes (see Table 6). The SRS data did not contain any outliers, but in all five of the outcome measures that had outliers to remove, improvement was observed. Of particular note was

the improvement in the question asking data. While this data had been non-significant previously, the removal of outliers resulted in a statistically significant outcome with a large effect size.

Table 6. *rANOVA Results with Outliers Removed*

	***p*-value**	**Cohen's *d***
Parent Surveys		
SMCS	<0.001	1.82[a]
SRS	na	na
Participant Surveys		
SMCS	0.003[a]	1.15[a]
SSIS	0.006[a]	1.08[a]
Behavioral Observations		
Percentage of Questions Asked	0.029[a]	0.95[a]
Percentage of Mutual Engagement	0.001	1.34[a]

a. Indicates a statistical improvement when outliers were removed from the data

Effect sizes. The effect size of each outcome variable, as seen in Tables 5 and 6, was calculated using the eta squared results from the rANOVA and converting to Cohen's *d*. Prior to the removal of outlying scores, five of the six outcome variables had large effect sizes and the sixth had a medium effect size. After the removal of outlying scores, all six variables indicated large treatment effects. The primary parent report measure, the SMCS Parent, was found to have a particularly large effect size ($d = 1.82$). Similarly, the primary participant report measure (the SMCS Participant) was also found to have a large effect size ($d = 1.15$). These results suggest that not only are participants who receive intervention improving more than those on a waitlist, but that participants are exhibiting very significant improvements during treatment as evaluated by a large variety of outcome measures.

Initial improvement. After assessing overall change during treatment, simple contrasts were evaluated to assess when significant change first occurred during the intervention in order to answer research question #2. Simple contrasts were obtained between Baseline scores and each subsequent time point (Week 5, Week 10, Week 15, and Week 20). Results were very similar between the full data set and the data set with outliers removed. The data with outliers removed is described below, with any deviations from the full data set explained. The primary parent report measure, the SMCS Parent, indicated significant improvement after 10 weeks of intervention, as did the primary participant report measure, the SMCS Participant. In addition, the SRS Parent and the question asking data also showed significant improvement after 10 weeks of intervention (although, as noted previously, when outliers were included the question asking data did not show any significant improvement). Importantly, the average SRS scores fell in the severe range at baseline and improved to the moderate range during treatment. The SSIS Participant measure did not show improvement until week 15 (when outliers were included, this change occurred at week 10). Interestingly, the mutual engagement data improved significantly after only 5 weeks of treatment. These data can be seen in Figure 1.

Continued improvement. Simple contrasts were again used to assess whether any significant change could be noted after the initial improvement point. When comparing later data points to the point of initial improvement (usually week 10), the primary parent report measure, the SMCS Parent, showed an additional statistically significant gain by week 20. No other outcome measures captured continued gains in social skills after the point of initial change. This data can be seen in Figure 1 as well.

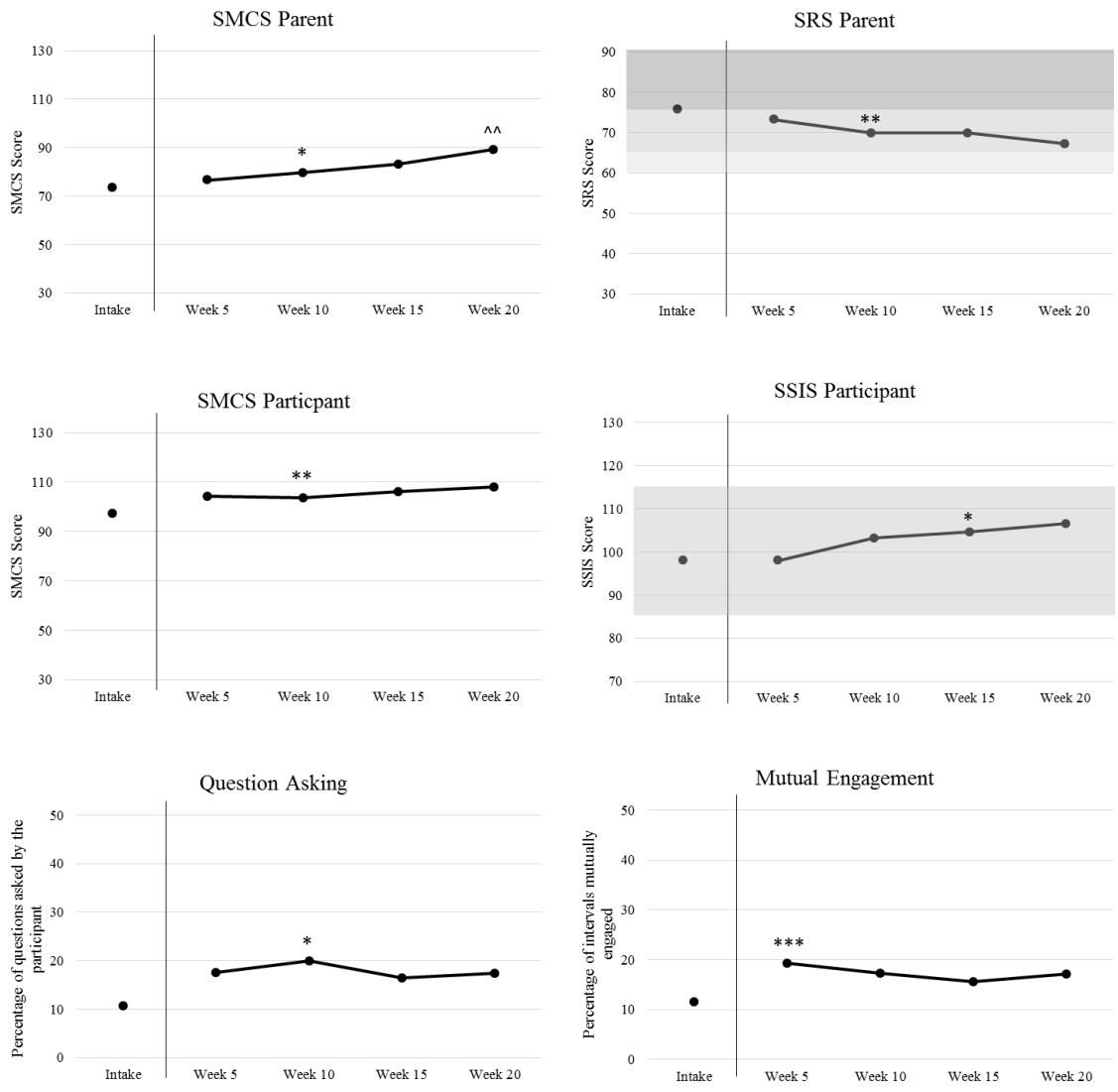

Figure 1. Graphs of repeated measures outcomes for each outcome variable with significant changes marked for both initial improvement over baseline and additional continued improvement. Initial improvement in scores is indicated as follows: * $p < .05$, ** $p < .01$, *** $p<.001$. Continued improvement in scores is indicated as follows: ^^ $p < .01$. On the SRS graph, the darkest band indicates scores in the severe range, the next darkest band indicates scores in the moderate range, and the lightest band indicates scores in the mild range. On the SSIS Participant graph, the grey band indicates scores that fall in the normative average range.

Discussion

The primary aims of this study were to assess the overall treatment effectiveness of the START group program when compared to a control group and to analyze the trajectory of those effects throughout the course of treatment. As a whole, the study was successful in finding preliminary evidence of the effectiveness of the START group as an intervention for high functioning adolescents with ASD. Because a number of analyses were run and each has its own implications, these analyses will be discussed in more detail below.

First, the study compared the final outcome measures of the treatment group with those of the waitlist control group. Effect sizes from this analysis were primarily medium to large and indicated clinically significant improvements in the treatment group above and beyond those of the waitlist group. While these results were not found to be statistically significant, there are a number of possible explanations for this non-significance. The extremely low levels of power obtained for this analysis point to an obvious explanation based on the number of participants enrolled at this stage in the trial. With additional participants, the improvement in power may very well lead to statistically significant results, especially when considering the large effect sizes observed in most of the outcome measures. An extension of this randomized controlled trial is already underway and these initial results suggest that improvement in significance will be highly likely when analyzing the data that is currently being collected, which will more than double the number of participants.

However, it is still prudent to consider other possible explanations for these non-significant results when comparing the treatment to a waitlist control. One explanation might be that the treatment is ineffective or less effective than anticipated. However, this

explanation is unlikely to be true once we consider the primarily large effect sizes obtained as well as the subsequent significant results of the repeated measures analysis. Another explanation might consider the fact that this study did find smaller effect sizes than might have been anticipated based on similar intervention studies and the results of a pilot study of the START group. Two published studies (using the same manualized intervention, PEERS) were identified that were most similar to the START group in terms of both the intervention (group trainings with parent and homework components) and age range (adolescents). These studies found effect sizes (Cohen's *d*) ranging from 0.8 to 1.45 (Laugeson et al., 2009; Laugeson, et al., 2012). Specifically examining the effect sizes found for the measures that are common to our two interventions results in effect sizes ranging from $d = 0.9$ to 1.37. As some of the outcome measures in the current study indicated effect sizes falling slightly below these, we could possibly conclude that there may be room for improvement in the START group that may result in larger effect sizes. One large difference between START and PEERS is that the PEERS intervention includes a concurrent parent group along with their adolescent treatment group. The inclusion of this additional component may result in greater gains and may be worthwhile to consider adding to the START group. On the other hand, the deeper parent involvement in the PEERS program may have resulted in greater bias in the parent outcome measures, leading to inflated effect sizes.

Another source of comparison lies in the pilot research done for the START program. With only six participants, the parent and adolescent report measures resulted in effect sizes as high as $d = 1.61$ and behavioral observations indicated effect sizes averaging $d = 1.55$. While several small changes were made to the intervention after the pilot study,

one noteworthy change was made that may conceivably have resulted in lower effect sizes. Throughout the duration of the pilot study, a licensed clinical psychologist and a graduate student facilitated the groups and met with participants and their parents. However, during the RCT, these tasks were assumed by undergraduate clinicians under professional supervision. It is plausible that this change in the intervention may have negatively impacted the observed effect sizes. However, because moderately large effect sizes were observed in the current study, which also utilized a vastly more rigorous research design, this change in treatment delivery is unlikely to be of great concern. In fact, this change was intentional, as the use of paraprofessionals as group facilitators is likely to make the treatment easier to implement in a variety of clinical, non-university settings and therefore more accessible to a large pool of clients who are in great need of appropriate and evidence-based interventions.

While the between-subjects analysis comparing the treatment group to a control group was non-significant, it is important to remember that due to the underpowered nature of the study, this result is to be expected at this preliminary stage of the trial. Therefore, it is more relevant to examine effect sizes when assessing the usefulness of the intervention, and here we find promising results suggesting that the treatment group is indeed exhibiting improvements above and beyond the control group with effect sizes ranging from small to large. In particular, the primary parent report measure as well as the primary participant report measure both improved considerably with large effect sizes of $d = 0.88$ and $d = 1.07$ respectively. These results are further supported by the moderate effect sizes noted in the objective behavioral observations data. Clearly, the START group is having a positive

impact on participants' social skills and this improvement can be observed by parents, by the participants themselves, as well as by naïve and blinded video coders.

Next, the study sought to examine the trajectory of these improvements. The within-subjects analysis comparing the baseline time point to later time points during treatment was significant in all measures (when outliers were removed), with all measures indicating particularly large effect sizes. While the between-subjects analysis indicated a treatment effect that surpassed any maturation of the waitlist group, the within-subjects analysis indicated that enormous improvements in social skills were made during the course of treatment, as assessed and supported by each of the various outcome measures. Simple contrasts revealed that significant improvement was first noted at week 10 for most of the measures. This finding is promising, as it indicates rapid improvement in symptoms and leaves room for the potential of continued improvement in the remaining ten weeks of the program. However, it was surprising to find that most measures did not appear to yield continued improvement at a statistically significant level. Although, the primary parent report measure did exhibit significant improvement in the remaining ten weeks of treatment and most of the other measures exhibited improving trends, albeit non-significant trends.

Again, there are a number of explanations for this observed trajectory that are worth considering. The simplest explanation might be that the START group is effective in improving social skills in the dosage of ten weeks, but that continued participation in the group does not result in continued gains. However, this explanation may be too simplistic and is at least partially called into question based on the continued improving trends in the data. Several other explanations may have some merit. First, it is possible that the inclusion of additional participants (as is currently underway) may increase statistical power and result

in the positive trends becoming statistically significant results. Second, it may be possible that we observe significant changes in the first ten weeks because those weeks are spent discussing and practicing the most basic social skills, which may be having the greatest amount of impact on the skills of the participants. Following this line of reasoning, it may be possible that the focus of the treatment in later weeks may be on topics that are too advanced for this population, resulting in a plateau of learning. If this were the case, future versions of the START group might consider revisiting the more basic topics in an attempt to continue to bolster participants' basic social skills. While this explanation may very well be true for some of the participants, it is likely that another explanation may also be important to consider.

The initial statistically significant improvement followed by improving but non-significant trends can also be explained by the nature of the outcome measurements employed in this study. Consider, for example, the behavioral observations data collected from five minute samples of a participant speaking with a stranger. In the first ten weeks of treatment the intervention addresses skills such as question asking, showing interest, and appropriate topics, all of which appeared to result in significant improvement in participants' ability to converse with a stranger. However, in the final ten weeks of treatment, the intervention addresses topics such as exhibiting good sportsmanship, giving and receiving feedback, working in a group, and using social media, all of which would be unlikely to be necessary skills to use in a brief conversation with a stranger. Therefore, it might be concluded that participants may indeed be learning new skills and experiencing social improvements in the final ten weeks of treatment, but that the outcome measures used in the current study were insufficiently able to capture this continued improvement.

While most of the outcome measures used in this study did appear to show responsiveness to the intervention, the initial approach to obtaining subjective ratings did not yield the anticipated results. After collecting the subjective rating data, it became clear that subjective ratings varied greatly between individual raters, resulting in substantial inter-rater variability. This variability appeared to have a greater impact on the raters' scores than did the participants' social skills, resulting in an ineffective outcome measure. In response, an alternative approach to obtaining subjective ratings was implemented, which controlled for this inter-rater variability. This subsequent approach resulted in substantial improvements and statistically significant differences were found when comparing the baseline to the final time point. This process represents a novel method for assessing the social significance of an intervention and these results are particularly exciting when considered in context. Not only do they provide additional evidence of the efficacy of the START program, they indicate that individuals who have no expertise in ASD, social interventions, or even the broader field of mental health, are clearly able to discern noticeable improvements in a participant's social skills after treatment. This exciting finding provides strong evidence for the true clinical and social significance of the intervention. Based on these promising findings, it is recommended that this line of research be continued to examine the amended subjective rating procedure results when comparing the treatment to the control group.

The positive findings from the behavioral observations data deserve further discussion. This data is particularly important due to the rigorous nature of collection. While other outcome measurements may be said to be biased in various ways, the method of collecting behavioral observations in the current study was rigorous, blinded, and objective, and therefore not subject to any potential biases. The large effect sizes noted in the within-

subjects portion of the study are therefore due particular attention, as they are likely representative of true behavioral change in participants' use of social skills. In fact, not only are the behavioral observations a rigorous outcome measure, they also provide some initial evidence of the generalization of skills to more naturalistic settings. As these observations examine a conversation that takes place between a participant and an unfamiliar peer, they allow us to assess for the use of learned social skills outside of the group and with an individual not associated with the group. Improvements in these behavioral observations therefore indicate that the participants are willing and able to put their new skills to use outside of the immediate setting of the treatment group.

When considered as a whole, the results from both phases of this study strongly suggest the efficacy of the START group as a treatment for adolescents with ASD. The participants' parents universally indicated extensive improvements in their child's social skills. Similarly, participants demonstrated insight into their own social skill improvements. Of particular importance, blinded and objective ratings of discrete social skills exhibited sizeable improvements, providing evidence for the generalization of these new skills outside of the group setting. Lastly, blinded subjective ratings of participants' social skills also showed improvement after the intervention, demonstrating the clinical and social significance of these improvements.

Research Design Considerations and Limitations

While the use of data from a randomized controlled trial allows for confidence in the research findings, several considerations still exist. The largest limitation that stands out above others is the use of the underpowered between-subjects analysis. Because the number of participants was too small to be able to detect statistical significance, it is quite likely that

the results were affected by Type II error, or an erroneous failure to reject the null hypothesis. However, this will soon be remedied by a substantial increase in the number of participants. In addition, despite the underpowered nature of the between-subjects analysis, the data did result in notable effect sizes, strongly suggesting social skill improvement due to the intervention. Lastly, the between-subjects analysis was followed by a within-subjects analysis which resulted in even larger effect sizes that were statistically significant, providing further evidence for the effectiveness of the START group and even further rationale for continuing the study with larger group sizes.

When interpreting the results from the within-subjects repeated-measures analysis, it is worth noting that a more rigorous within-subjects design would involve counterbalancing the order of topics presented within the intervention. This was not possible to implement in the current study due to the logically progressive nature of the intervention. As noted earlier, Krasny, Williams, Provencal, and Ozonoff (2003) called for social interventions that "program in a sequential and progressive manner." Introducing more complex social skills prior to introducing basic social skills would have been highly likely to diminish the helpfulness of the intervention, therefore, a counterbalanced design was not practical to implement. However, interpretations of the within-subjects results can be strengthened by the understanding that this data comes from an RCT, which has established the usefulness of the intervention above and beyond a waitlist control group.

One threat to the internal validity of the study must also be acknowledged as a limitation. Testing, or the practice effect, may be of concern in the current study, as the outcome measures are administered repeatedly. While the waitlist control group partially controls for this threat to internal validity, it is possible that the waitlist group may not have

been as affected by practice effects as the treatment group. The waitlist control group completed one pre-test, a 20-week wait, and one post-test. However, the treatment group was given assessments at five week intervals and therefore completed four tests before their final post-test. In particular, this could affect the behavioral outcome measures, as participants in the treatment group had three more chances to "practice" having conversations than the participants in the waitlist group. In the future, it is recommended that the START group be studied with another RCT which does not involve repeatedly collecting data every five weeks in order to control for any possible testing effects. However, this may not be necessary, as parent and participant report questionnaires are not subject to practice effects, and similar improvements were noted between the parent report, participant report, and behavioral observation data, indicating that the improvements observed must not have been significantly impacted by a testing effect.

Instrumentation, or changes in the measurement over the course of a study, could be argued to be another potential threat to the internal validity of this study, particularly in regard to the videotaped behavioral observations. These behavioral observations were made by coders, and it is known that coders often "drift" in their coding over time (Kazdin, 1982). However, this threat was controlled for in several ways. First, the videos were presented to the coders in random order in order to ensure that any drift would occur randomly throughout the observations rather than systematically toward the end of the intervention. In addition, multiple raters coded each video and established interrater reliability, as described above in the methods. Interrater reliability does not completely guarantee a control for observer drift, as both observers could theoretically drift at the same rate, but it does provide a strong assurance that major deviations from the original coding scheme did not occur.

The external validity of the study may be an additional minor limitation. Generalizability across people can only be made statistically possible by using true random sampling from an entire population of interest. However, because true (or even approximate) random sampling is practically impossible in intervention research, generalization in the statistical sense cannot be made and we must rely on "rational generalization" (Heppner, Wampold, & Kivlighan, 2008). This requires primarily that the researcher establish the characteristics of the participants and limit their generalizations to populations with those characteristics. External validity is typically fairly robust in applied research (Heppner, Wampold, & Kivlighan, 2008), as real clients who are seeking help are being treated in settings where real clients would typically be seeking help. In the case of the current study, a much-needed service was offered, it was advertised to the local community, and the first 20 clients who contacted us with interest (and met inclusion criteria) were accepted as participants in the study. This is the exact method by which a real-world treatment would typically be implemented. Especially given the relatively few exclusion criteria for participation in the study, it is "rational" to anticipate that the results of this study could be at least generalizable to other centers that treat adolescents with ASD, especially if they are located in communities similar to Santa Barbara (ethnically, geographically, socioeconomically, etc.). Generalizability to broader populations will need to be examined with future research. Additionally, as social interactions can vary significantly across cultures, culturally relevant adaptations of the START group may be useful when implementing the intervention with other populations.

The lack of a confirmatory ASD diagnosis is another limitation of the current study. It is possible, though unlikely, that participants in the study may not have been diagnosed

with an ASD according to strict DSM5 criteria after a comprehensive assessment. Future research assessing the START group may consider performing a more comprehensive diagnostic evaluation in order to determine inclusion in the study. However, the current study did require a previous ASD diagnosis as well as a score above 60 on the SRS (which is the normed cutoff for the presence of an ASD), which provided some degree of confidence in the diagnostic status of participants.

Future Directions

The findings of the current study provide inspiration for many future research directions. In addition to the expansion of the current project to increase the number of participants and other suggestions made above, several further areas for future research present themselves.

While the current findings provide evidence of the positive impact the START group has on participants' social skills, the maintenance of treatment gains after completion of the intervention has not yet been addressed. As the ultimate goal of intervention is always to promote lasting change, this will be an important area to study further. In fact, six month follow-up data is already in the process of being collected for the participants in this study and it is hoped that this data will shed some light on whether or not the START group does in fact have a lasting positive impact.

While the current study assessed the usefulness of the START program in a clinical setting, its usefulness in other settings may also be important to understand. For example, adaptation to the school setting would allow for even greater distribution of the treatment, but may require some substantial modification of the current procedures in order to adapt to the time and resource constraints of the school setting. The START group was intentionally

designed with wide distribution of the protocol in mind, so it is likely that implementing the program in other clinical settings, whether university associated or not, should require relatively little modification.

Additionally, it is recommended that future research consider possible outcome predictors for the START group. It was noted in the current study that a great deal of variability in baseline characteristics and response to intervention existed between participants. In order to maximize the treatment effect it may be helpful to better understand factors that may predict superior outcomes. Possible predictors could be Verbal IQ level, motivation to engage in treatment, social motivation, or comorbid diagnoses, to name only a few.

The subjective rating procedure initially used in this study was clearly unable to capture the improvements in social skills that data from other sources indicated. However, the endeavor to obtain social significance ratings is an important one and warrants further exploration of potential avenues for improvement. Already, vast improvement was made by modifying the procedure to reduce inter-rater variability. It is suggested that this alternative method for collecting subjective ratings be used in the future to examine social skills improvement in the current study while comparing the treatment group to the waitlist control. It is also recommended that the subjective rating findings be compared to a number of objective behavioral observations in order to elucidate the possible concrete behaviors that may be subjectively assessed by naïve raters. This endeavor could help to further our understanding of what specific behaviors give the impression of social competence, important information for the development and improvement of both the START group and any other social skill intervention.

The current study assessed two objective behavioral measures: question asking and mutual engagement. Many other social skills are likely being captured in the videotaped conversations that may lend themselves to behavioral coding and statistical analysis, and continued assessment of these additional behavioral measures may provide an exciting area of future research. Further exploration of these skills will again help to refine our understanding of what specifically is changing during the course of this intervention, as well as how much that skill correlates with subjective rating improvements.

As discussed above, there is a high likelihood that participants in the START group continue to receive benefit from the intervention after the first ten weeks, yet our current outcome measures were not sufficiently able to capture this change. To address this problem, it is suggested that future research studies attempt to develop supplemental measures to assess complex social change in interventions such as the START group. These measures would likely be useful not only for studying outcomes in social skills groups for adolescents with ASD, but in any intervention research targeting social skills across various diagnoses and across the lifespan.

Lastly, future studies of the START group may consider the inclusion of additional rigorous outcome measurements. Possible examples include teacher reports (as teachers can be kept blind to the treatment status of the participant), blinded clinician ratings of social skills and ASD symptomatology, or behavioral observations in more natural settings. Similarly, it may be useful to conduct behavioral observations of typical peers in order to better understand the normative range of behaviors and thereby assist in setting appropriate behavioral goals. While these supplemental measures have the potential to provide additional evidence for the efficacy of the START group, the current study did implement a

number of rigorous outcome measurements, exceeding the norm for social skills interventions. It is recommended that future research of other social interventions also strive toward the use of un-biased measurement tools.

Implications

The current study addressed two important research questions about the efficacy of the START group and the trajectory of improvement in participants. The positive findings of this initial research on the START group provide strong preliminary evidence for the clinical effectiveness of this treatment approach. These results also provide a rationale for continuing to assess the effectiveness of the START group, particularly through increasing the number of participants in the current study.

In addition, these results have provided some food for thought on the optimal length of treatment using the START group. While the length of the intervention is currently set at 20 weekly sessions following a progressively more complex curriculum, findings from this study may inform future changes to this protocol. Several possible directions present themselves based on the data. Firstly, persisting with the current protocol is certainly supported by the large effect sizes found and the continued improving trends in the data throughout the entirety of the intervention. The new data currently being gathered from additional participants will continue the current line of research and will provide further insight into the trajectory of improvement under the existing protocol. Secondly, it may be worthwhile to consider the addition of supplementary outcome measurements that are tailored toward assessing the more complicated types of social change that may be occurring later in the intervention, as discussed above. As the primary parent report measure indicated that parents were observing continued statistically significant change throughout the entire

course of treatment, it is likely that supplementary outcome measurements may be able to provide further insight into the types of changes that are occurring in the final ten weeks of treatment. Thirdly, and only if the addition of supplementary outcome measurements does not result in the observation of continued social skill improvements after the first ten weeks, it may be worth considering amending the curriculum to continue to focus on basic social skill topics even after the first ten weeks. One way of achieving this might be to create a modular approach to the START curriculum with both basic and advanced modules. In this way, participants could have the ability to repeat the basic module as many times as necessary in order to bolster their use of basic social skills, while advancing to a more sophisticated social skill module only after meeting certain basic benchmarks. This format would also allow for further individualization of the intervention, a factor that (as discussed in the literature review) may provide some benefit when working with a diverse range of participants.

Regardless of the future course of research for the START group, this examination of the trajectory of improvement of participants has satisfied an important need in the social skill group intervention literature. As noted in the literature review, no studies have previously examined length of treatment and treatment trajectories, and the current study provides an essential first step in the investigation of this key aspect of understanding an intervention. While the findings of the current study are specific to the START group, it is possible that they may also be of some use in informing the optimal length of other social skills groups for adolescents with ASD as well.

While the current study provides evidence of the effectiveness of the START group, it also attests to something even more basic: the plasticity of social skills. Teens with ASD

may lag behind their peers in social understanding, motivation, and/or skill implementation, but studies like these provide hope for the possibility of great improvement in these areas. These deficits are not permanent or intractable but are incredibly malleable, even after only ten sessions of intervention.

As we are all aware, adolescence can be a particularly difficult time in the life of any individual. Emotions run high, independence is sought from one's parents, and social demands become more and more complex. Adolescents are expected to make friends without the assistance of their parents setting up play dates, maintain and develop those friendships over time, and navigate a host of tricky relational situations such as teasing, bullying, flirting, joking, texting, and dating. Even the most socially competent teens are likely to experience some social distress during this period (didn't you?). Yet, as discussed above, adolescence has been a historically understudied area in the ASD literature, and few interventions for teens have been studied rigorously and systematically (Reichow & Volkmar, 2010).

In fact, adolescence represents a critical period for intervention more reasons than just the complexity of this developmental period. It is known that the symptoms of ASD persist throughout this developmental era, even with early intervention (Schall & McDonough, 2010; Seltzer et al., 2003; Sigman & Ruskin, 1999). Further, adolescence may be the last point in time during which treatment providers can leverage the motivation of parents. Not only are parents instrumental in seeking out and pursuing treatment opportunities for their children, they can also be active allies in the process of intervention and particularly in promotion of the generalization of skills outside of the treatment setting. The START group and others like it are able to capitalize on parents' motivation by

involving them in treatment and giving them tools to help their teens. This can be particularly helpful in the generalization and maintenance of treatment gains, as parents are likely to have the opportunity to encourage social skill use in other settings and beyond the duration of the intervention.

By intervening at this crucial developmental stage, we have the opportunity to alter the social trajectory of these teens. Past research has discovered that without the appropriate skill set to negotiate social situations, adolescents with ASD can unfortunately fall victim to a host of unpleasant consequences. Bullying and victimization rates are much higher for individuals on the autism spectrum in comparison with their typically developing peers (Carter, 2009; Little, 2001; van Roekel, Scholte, & Didden, 2010). In fact, Little's study found that victimization rates were four times higher in adolescents with ASD, with up to 75% of this population reporting victimization. In addition, adolescence is a time when many individuals with ASD begin to experience the devastating effects of other comorbid psychiatric diagnoses, such as depression and anxiety (Church, Alisanski, & Amanullah, 2000; Ghaziuddin, Ghaziuddin, & Greden, 2002; Kim, Szatmari, Bryson, Streiner, & Wilson, 2000). A recent study found that 70% of adolescents with ASD had symptoms consistent with at least one comorbid disorder and 41% had two or more comorbid diagnoses (Simonoff et al., 2008). The most commonly reported comorbid diagnosis across a number of studies is anxiety. As friendships have consistently been found to be a protective factor against mental health issues (Mazurek, 2014; Miller & Ingham, 1976), it is reasonable to assume that by helping adolescents with ASD to develop friendships, interventions like the START group may be able to improve not only their ASD symptoms, but their comorbid symptoms as well.

As many researchers have noted, autism interventions may have the ability to correct an atypical developmental course (e.g. Gulsrud, Hellemann, Freeman, & Kasari, 2014; Klintwall, Eldevik, & Eikeseth, 2015). Without intervention, these adolescents would at a minimum experience significant social struggles. Their impaired social awareness and skill use could lead to a lack of friendships, falling victim to bullying, or becoming susceptible to depression or anxiety. Conceivably, they could over time become discouraged from even attempting social interactions or lose interest in social interactions entirely, which is sometimes reported by adults with ASD (Baron-Cohen & Wheelwright, 2003). However, social skills intervention has the potential to prevent these morose outcomes, allowing for greater possibilities for social success. After only ten weeks of intervention, participants in the current study reported statistically significant increases in their confidence in their social skills, *and* unfamiliar raters also observed an increase in social confidence. Even if their skill use remained unchanged (which it doesn't, based on the behavioral observations and subjective rating data), this increase in confidence alone would bode very well for the social trajectory of participants.

In fact, altering an adolescent's social trajectory may collaterally alter other important areas of their lives as well. While more research is necessary to determine the long-term outcomes of intervention participants, it is quite possible that social improvements in the teenage years could impact their ability to obtain employment, to find a romantic partner, and to otherwise allow them to reach the many personal goals that are facilitated by social acumen.

In summary, by providing a necessary intervention at a crucial developmental time, the START group is able to effectively improve the social skills of participants, which may

in turn alter their developmental course, leading to broadly improved outcomes across their lives. Further research will be necessary to understand more fully both the immediate and the far-reaching implications of this intervention, but this study provides the essential first steps toward understanding the immediate impact this intervention has on improving the social skills of adolescents with ASD.

References

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (5th Ed.)*. Arlington, VA: American Psychiatric Publishing.

Barnhill, G. P., Tapscott Cook, K., Tebbenkamp, K., & Smith Myles, B. (2002). The Effectiveness of Social Skills Intervention Targeting Nonverbal Communication for Adolescents With Asperger Syndrome and Related Pervasive Developmental Delays. *Focus on Autism and Other Developmental Disabilities, 17*(2), 112-118. doi: http://dx.doi.org/10.1177/10883576020170020601

Baron-Cohen, S., & Wheelwright, S. (2003). The friendship questionnaire: An investigation of adults with asperger syndrome or high-functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 33*(5), 509-517. doi:http://dx.doi.org/10.1023/A:1025879411971

Barry, T. D., Klinger, L. G., Lee, J. M., Palardy, N., Gilmore, T., & Bodin, S. D. (2003). Examining the effectiveness of an outpatient clinic-based social skills group for high-functioning children with autism. *Journal of Autism and Developmental Disorders, 33*, 685–701.

Begeer, S., Gevers, C., Clifford, P., Verhoeve, M., Kat, K., Hoddenbach, E., & Boer, F. (2010). Theory of mind training in children with autism: A randomized controlled trial. Journal of Autism and Developmental Disorders, 41(8), 997-1006. http://dx.doi.org/10.1007/s10803-010-1121-9

Broderick, C., Caswell, R., Gregory, S., Marzolini, S., & Wilson, O. (2002). 'Can I join the Club?'. A Social Integration Scheme for Adolescents with Asperger Syndrome. *Autism, 6*(4), 427-431. doi: http://dx.doi.org/10.1177/1362361302006004008

Carter, C., Meckes, L., Pritchard, L., Swensen, S., Wittman, P. P., & Velde, B. (2004). The friendship club: An after-school program for children with Asperger syndrome. *Family & Community Health: The Journal of Health Promotion & Maintenance, 27*(2), 143-150.

Carter, S. (2009). Bullying of students with Asperger syndrome. *Issues in Comprehensive Pediatric Nursing, 32*(3), 145-154. doi: http://dx.doi.org/10.1080/01460860903062782

Centers for Disease Control and Prevention (2014). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2010. *Morbidity and Mortality Weekly Report Surveillance Summaries, 63*, 1-21.

Church, C., Alisanski, S., & Amanullah, S. (2000). The social, behavioral, and academic experiences of children with asperger syndrome. *Focus on Autism and Other Developmental Disabilities, 15*(1), 12-20. doi:10.1177/108835760001500102

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Constantino, J. N. & Gruber, C. P. (2005) *Social Responsiveness Scale (SRS)*. Western Psychological Services.

Corbett, B. A., Gunther, J. R., Comins, D., Price, J., Ryan, N., Simon, D., Schupp, C. W., & Rios, T. (2011). Brief Report: Theatre as Therapy for Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders, 41*(4), 505-511.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions

about growth curve modeling. *Journal of Cognition and Development, 11*(2), 121-136. doi:http://dx.doi.org/10.1080/15248371003699969

Dotson, W. H., Leaf, J. B., Sheldon, J. B., & Sherman, J. A. (2010). Group teaching of conversational skills to adolescents on the autism spectrum. *Research in Autism Spectrum Disorders, 4*(2), 199-209. doi: http://dx.doi.org/10.1016/j.rasd.2009.09.005

Epp, K. M. (2008). Outcome-Based Evaluation of a Social Skills Program Using Art Therapy and Group Therapy for Children on the Autism Spectrum. *Children & Schools, 30*(1), 27-36.

Field, A. (1998). A bluffer's guide to ... sphericity. *The British Psychological Society: Mathematical, Statistical & Computing Section Newsletter, 6*, 13-22.

Frankel, F., Myatt, R., Sugar, C., Whitham, C., Gorospe, C. M., & Laugeson, E. (2010). A Randomized Controlled Study of Parent-assisted Children's Friendship Training with Children having Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 40*(7), 827-842. doi: 10.1007/s10803-009-0932-z

Fullerton, A., & Coyne, P. (1999). Developing Skills and Concepts for Self-Determination in Young Adults with Autism. *Focus on Autism and Other Developmental Disabilities, 14*(1), 42-52. doi: http://dx.doi.org/10.1177/108835769901400106

Gabriels, R. L., Agnew, J. A., Holt, K. D., Shoffner, A., Zhaoxing, P., Ruzzano, S., Clayton, G. H., & Mesibov, G. (2012). Pilot study measuring the effects of therapeutic horseback riding on school-age children and adolescents with autism spectrum disorders. *Research in Autism Spectrum Disorders, 6*(2), 578-588. doi: http://dx.doi.org/10.1016/j.rasd.2011.09.007

Ghaziuddin, M., Ghaziuddin, N., & Greden, J. (2002). Depression in persons with

autism: Implications for research and clinical care. *Journal of Autism and Developmental Disorders, 32*(4), 299-306. doi: 10.1023/A:1016330802348

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research, 42*(3), 237-288. doi:http://dx.doi.org/10.3102/00346543042003237

Green, J., Gilchrist, A., Burton, D., & Cox, A. (2000). Social and psychiatric functioning in adolescents with asperger syndrome compared with conduct disorder. *Journal of Autism and Developmental Disorders, 30*(4), 279-293. doi: 10.1023/A:1005523232106

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95-112. doi:http://dx.doi.org/10.1007/BF02289823

Gresham, F.M., & Elliott, S.N. (1990). Social Skills Rating System manual. Circle Pines, MN: AGS.

Gulsrud, A. C., Hellemann, G. S., Freeman, S. F. N., & Kasari, C. (2014). Two to ten years: Developmental trajectories of joint attention in children with ASD who received targeted social communication interventions. *Autism Research, 7*(2), 207-215. doi:http://dx.doi.org/10.1002/aur.1360

Heppner, P. P., Kivlighan, D. M., Jr., & Wampold, B. E. (1992). *Research design in counseling.* Thomson Brooks/Cole Publishing Co, Belmont, CA.

Herbrecht, E., Poustka, F., Birnkammer, S., Duketis, E., Schlitt, S., Schmoetzer, G., & Boelte, S. (2009). Pilot evaluation of the Frankfurt Social Skills Training for children

and adolescents with autism spectrum disorder. *European Child & Adolescent Psychiatry, 18*(6), 327-335. doi: http://dx.doi.org/10.1007/s00787-008-0734-4

Hillier, A., Greher, G., Poto, N., & Dougherty, M. (2012). Positive outcomes following participation in a music intervention for adolescents and young adults on the autism spectrum. *Psychology of Music, 40*(2), 201-215. doi: http://dx.doi.org/10.1177/0305735610386837

Huynh, H., & Feldt, L. S. (1976). Estimation of the box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics,* 1.1, 69-82.

Kazdin, A. E. (1982). Single-case experimental designs in clinical research and practice. *New Directions for Methodology of Social & Behavioral Science, 13,* 33-47.

Kempe, A., & Tissot, C. (2012). The use of drama to teach social skills in a special school setting for students with autism. *Support for Learning*, 27(3), 97-102.

Kim, J. A., Szatmari, P., Bryson, S. E., Streiner, D. L., & Wilson, F. J. (2000). The prevalence of anxiety and mood problems among children with autism and asperger syndrome. *Autism, 4*(2), 117-132. doi: 10.1177/1362361300004002002

Klintwall, L., Eldevik, S., & Eikeseth, S. (2015). Narrowing the gap: Effects of intervention on developmental trajectories in autism. *Autism, 19*(1), 53-63. doi:http://dx.doi.org/10.1177/1362361313510067

Koegel, L. K., Vernon, T. W., Koegel, R. L., Koegel, B. L., & Paulin, A. W. (2012). Improving social engagement and initiations between children with autism spectrum disorder and their peers in inclusive settings. Journal of Positive Behavior Support, 14(4), 220-227. http://dx.doi.org/10.1177/1098300712437042

Krasny, L., Williams, B. J., Provencal, S., & Ozonoff, S. (2003). Social skills interventions for the autism spectrum: Essential ingredients and a model curriculum. *Child and Adolescent Psychiatric Clinics of North America, 12*(1), 107-122. doi:http://dx.doi.org/10.1016/S1056-4993(02)00051-2

Laugeson, E. A., Frankel, F., Gantman, A., Dillon, A. R., & Mogil, C. (2012). Evidence-Based Social Skills Training for Adolescents with Autism Spectrum Disorders: The UCLA PEERS Program. *Journal of Autism and Developmental Disorders, 42*(6), 1025-1036. doi: http://dx.doi.org/10.1007/s10803-011-1339-1

Laugeson, E. A., Frankel, F., Mogil, C., & Dillon, A. R. (2009). Parent-Assisted Social Skills Training to Improve Friendships in Teens with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 39*(4), 596-606. doi: http://dx.doi.org/10.1007/s10803-008-0664-5Little, L. (2001). Peer victimization of children with Asperger spectrum disorders. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(9), 995-996. doi: http://dx.doi.org/10.1097/00004583-200109000-00007

Lerner, M. D., & Mikami, A. Y. (2012). A Preliminary Randomized Controlled Trial of Two Social Skills Interventions for Youth With High-Functioning Autism Spectrum Disorders. *Focus on Autism and Other Developmental Disabilities, 27*(3), 147-157. doi: http://dx.doi.org/10.1177/1088357612450613

Liu, S. (2008). Comparing repeated-measures ANOVA and multilevel growth curve modeling: A simulation study. Thesis - Pennsylvania State University, 2008. http://etda.libraries.psu.edu/theses/approved/WorldWideIndex/ETD-3430/index.html.

Lopata, C., Thomeer, M. L., Volker, M. A., Nida, R. E., & Lee, G. K. (2008). Effectiveness of a Manualized Summer Social Treatment Program for High-Functioning Children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 38*(5), 890-904. doi: http://dx.doi.org/10.1007/s10803-007-0460-7

MacKay, T., Knott, F., & Dunlop, A.-W. (2007). Developing social interaction and understanding in individuals with autism spectrum disorder: A groupwork intervention. *Journal of Intellectual and Developmental Disability, 32*(4), 279-290. doi: http://dx.doi.org/10.1080/13668250701689280

Mazurek, M. O. (2014). Loneliness, friendship, and well-being in adults with autism spectrum disorders. *Autism, 18*(3), 223-232. doi:http://dx.doi.org/10.1177/1362361312474121

McMahon, C. M., Vismara, L. A., & Solomon, M. (2013). Measuring changes in social behavior during a social skills intervention for higher-functioning children and adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 43*(8), 1843-1856. doi: http://dx.doi.org/10.1007/s10803-012-1733-3

Mesibov, G. B. (1984). Social skills training with verbal autistic adolescents and adults: A program model. *Journal of Autism and Developmental Disorders, 14*(4), 395-404. doi: http://dx.doi.org/10.1007/BF02409830

Miller, A. R., Vernon, T. W., Wu, V., & Russo, K. (2014). Social skill group interventions for adolescents with autism spectrum disorders: A systematic review. *Review Journal of Autism and Developmental Disorders.* doi: 10.1007/s40489-014-0017-6

Miller, P. M., & Ingham, J. G. (1976). Friends, confidants and symptoms. *Social Psychiatry, 11*(2), 51-58. doi: http://dx.doi.org/10.1007/BF00578738

Minihan, A., Kinsella, W., & Honan, R. (2011). Social skills training for adolescents with Asperger's syndrome using a consultation model. *Journal of Research in Special Educational Needs, 11*(1), 55-69. doi: http://dx.doi.org/10.1111/j.1471-3802.2010.01176.x

Mishna, F., & Muskat, B. (1998). Group Therapy for Boys with Features of Asperger Syndrome and Concurrent Learning Disabilities: Finding a Peer Group. *Journal of Child and Adolescent Group Therapy, 8*(3), 97-114. doi: 10.1023/A:1022984118001

Mitchell, K., Regehr, K., Reaume, J., & Feldman, M. (2010). Group social skills training for adolescents with Asperger Syndrome or high functioning autism. *Journal on Developmental Disabilities, 16*(2), 52-63.

National Autism Center (2009). National Standards Report. Randolph, MA: National Autism Center

Ozonoff, S., & Miller, J. N. (1995). Teaching Theory of Mind: A New Approach to Social Skills Training for Individuals with Autism. *Journal of Autism and Developmental Disorders, 25*(4), 415-433.

Reichow, B., Steiner, A. M., & Volkmar, F. (2012). Cochrane Review: Social skills groups for people aged 6 to 21 with autism spectrum disorders (ASD). *Evidence-Based Child Health: A Cochrane Review Journal, 8*(2), 266-315. doi: 10.1002/ebch.1903

Reichow, B., & Volkmar, F. R. (2010). Social skills interventions for individuals with autism: Evaluation for evidence-based practices within a best evidence synthesis

framework. *Journal of Autism and Developmental Disorders, 40*(2), 149-166. doi: http://dx.doi.org/10.1007/s10803-009-0842-0

Rose, R., & Anketell, C. (2009). The Benefits of Social Skills Groups for Young People with Autism Spectrum Disorder: A Pilot Study. *Child Care in Practice, 15*(2), 127-144. doi: http://dx.doi.org/10.1080/13575270802685377

Schall, C. M., & McDonough, J. T. (2010). Autism spectrum disorders in adolescence and early adulthood: Characteristics and issues. *Journal of Vocational Rehabilitation, 32*(2), 81-88.

Seltzer, M. M., Krauss, M. W., Shattuck, P. T., Orsmond, G., Swe, A., & Lord, C. (2003). The symptoms of autism spectrum disorders in adolescence and adulthood. *Journal of Autism and Developmental Disorders, 33*(6), 565-581. doi: http://dx.doi.org/10.1023/B:JADD.0000005995.02453.0b

Shin, S., Koh, M. S., & Yeo, M. H. (2012). A Comparative Study of the Preliminary Effects in the Levels of Adaptive Behaviors: Learning Program for the Development of Children with Autism (LPDCA). *Journal of the International Association of Special Education, 13*(1), 6-15.

Sigman, M., & Ruskin, E. (1999). Continuity and change in the social competence of children with autism, Down syndrome, and developmental delays. *Monographs of the Society for Research in Child Development, 64*(1), 1-139.

Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: Prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of the*

American Academy of Child & Adolescent Psychiatry, *47*(8), 921-929. doi: http://dx.doi.org/10.1097/CHI.0b013e318179964f

Simpson, R. L., de Boer-Ott, S. R., Griswold, D. E., Myles, B. S., Byrd, S. E., Ganz, J. B., Tapscott Cook, K., Otten, K. L., Ben-Arieh, J., Kline, S. A., & Adams, L. G. (2005). *Autism spectrum disorders: Interventions and treatments for children and youth*: Corwin Press, Thousand Oaks, CA.

Smith, T., Scahill, L., Dawson, G., Guthrie, D., Lord, C., Odom, S., Rogers, S., & Wagner, A. (2007). Designing research studies on psychosocial interventions in autism. *Journal of Autism and Developmental Disorders, 37*(2), 354-366. doi: http://dx.doi.org/10.1007/s10803-006-0173-3

Stevens, J. (1992). *Applied multivariate statistics for the social sciences (2nd ed.).* Lawrence Erlbaum Associates, Inc, Hillsdale, NJ.

Stichter, J. P., Herzog, M. J., O'Connor, K. V., & Schmidt, C. (2012). A Preliminary Examination of a General Social Outcome Measure. *Assessment for Effective Intervention, 38*(1), 40-52. doi: http://dx.doi.org/10.1177/1534508412455213

Stichter, J. P., Herzog, M. J., Visovsky, K., Schmidt, C., Randolph, J., Schultz, T., & Gage, N. (2010). Social Competence Intervention for Youth with Asperger Syndrome and High-functioning Autism: An Initial Investigation. *Journal of Autism and Developmental Disorders, 40*(9), 1067-1079. doi: http://dx.doi.org/10.1007/s10803-010-0959-1

Stokes T. F, & Baer D. M. An implicit technology of generalization. *Journal of Applied Behavior Analysis*. (1977);10:349–367. doi:10.1901/jaba.1977.10-349.

Tse, J., Strulovitch, J., Tagalakis, V., Meng, L., & Fombonne, E. (2007). Social Skills Training for Adolescents with Asperger Syndrome and High-Functioning Autism. *Journal of Autism and Developmental Disorders, 37*(10), 1960-1968. doi: http://dx.doi.org/10.1007/s10803-006-0343-3

Tyminski, R. F., & Moore, P. J. (2008). The impact of group psychotherapy on social development in children with pervasive development disorders. *International Journal of Group Psychotherapy, 58*(3), 363-379. doi: http://dx.doi.org/10.1521/ijgp.2008.58.3.363

van Roekel, E., Scholte, R. H. J., & Didden, R. (2010). Bullying among adolescents with autism spectrum disorders: Prevalence and perception. *Journal of Autism and Developmental Disorders, 40*(1), 63-73. doi: http://dx.doi.org/10.1007/s10803-009-0832-2

Valenti, M., Cerbo, R., Masedu, F., De Caris, M., & Sorge, G. (2010). Intensive intervention for children and adolescents with autism in a community setting in Italy: A single-group longitudinal study. *Child and Adolescent Psychiatry and Mental Health, 4*. doi: http://dx.doi.org/10.1186/1753-2000-4-23

Watkins, L., O'Reilly, M., Kuhn, M., Gevarter, C., Lancioni, G. E., Sigafoos, J., & Lang, R. (2015). A review of peer-mediated social interaction interventions for students with autism in inclusive settings. *Journal of Autism and Developmental Disorders,45*(4), 1070-1083. doi:http://dx.doi.org/10.1007/s10803-014-2264-x

Webb, B. J., Miller, S. P., Pierce, T. B., Strawser, S., & Jones, W. (2004). Effects of Social Skill Instruction for High-Functioning Adolescents With Autism Spectrum

Disorders. *Focus on Autism and Other Developmental Disabilities, 19*(1), 53-62. doi: http://dx.doi.org/10.1177/10883576040190010701

White, S. W., Albano, A. M., Johnson, C. R., Kasari, C., Ollendick, T., Klin, A., Oswald, D., & Scahill, L. (2010). Development of a Cognitive-Behavioral Intervention Program to Treat Anxiety and Social Deficits in Teens with High-Functioning Autism. *Clinical Child and Family Psychology Review, 13*(1), 77-90.

White, S. W., Koenig, K., & Scahill, L. (2010). Group Social Skills Instruction for Adolescents with High-Functioning Autism Spectrum Disorders. *Focus on Autism and Other Developmental Disabilities, 25*(4), 209-219.

White, S. W., Ollendick, T., Albano, A. M., Oswald, D., Johnson, C., Southam-Gerow, M. A., Kim, I., & Scahill, L. (2013). Randomized controlled trial: Multimodal anxiety and social skill intervention for adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 43*(2), 382-394. doi: http://dx.doi.org/10.1007/s10803-012-1577-x

White, S. W., Ollendick, T., Scahill, L., Oswald, D., & Albano, A. M. (2009). Preliminary Efficacy of a Cognitive-Behavioral Treatment Program for Anxious Youth with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 39*(12), 1652-1662. doi: http://dx.doi.org/10.1007/s10803-009-0801-9

Williams, T. I. (1989). A social skills group for autistic children. *Journal of Autism and Developmental Disorders, 19*(1), 143-155. doi: http://dx.doi.org/10.1007/BF02212726

Williams White, S., Keonig, K., & Scahill, L. (2007). Social skills development in

children with autism spectrum disorders: A review of the intervention research. *Journal of Autism and Developmental Disorders, 37*(10), 1858-1868. doi: 10.1007/s10803-006-0320-x

Wright, C., Diener, M. L., Dunn, L., Wright, S. D., Linnell, L., Newbold, K., D'Astous, V., & Rafferty, D. (2011). SketchUpTM: A Technology Tool to Facilitate Intergenerational Family Relationships for Children with Autism Spectrum Disorders (ASD). *Family and Consumer Sciences Research Journal, 40*(2), 135-149. doi: http://dx.doi.org/10.1111/j.1552-3934.2011.02100.x

Zhang, J., & Wheeler, J. J. (2011). A meta-analysis of peer-mediated interventions for young children with autism spectrum disorders. *Education and Training in Autism and Developmental Disabilities, 46*(1), 62-77.

Appendix A

START Weekly Topics

Week	Topic
1	Greeting others/ making initial introductions
2	Using questions to maintain a conversation
3	Using comments to maintain a conversation
4	Showing interest – attention, eye contact, facial expressions
5	Choosing appropriate topics for conversations
6	Making and keeping friends
7	Changing topics, ending conversations, and saying goodbye
8	Reducing anxiety/being comfortable during social exchanges
9	Expressing empathy
10	Complimenting others
11	Respectfully disagreeing with others
12	Demonstrating good sportsmanship/being a good winner and loser
13	Making a good impression/giving social feedback
14	Receiving social feedback
15	Working in a group/being a good team member/leader
16	Using appropriate humor and understanding/using sarcasm
17	Having social courage/joining a new group of peers
18	Using social media
19	Hosting others at your home/being a good guest at another person's home
20	Summary and conclusion